Thèse présentée pour obtenir le titre de

DOCTEUR DE L'ECOLE NATIONALE DES PONTS ET CHAUSSEES

Spécialité : Mathématiques appliquées

par

Jérôme LELONG

Etude asymptotique des algorithmes stochastiques et calcul du prix des options Parisiennes

Thèse soutenue le 14 septembre 2007 devant le jury composé de :

Monique JEANBLANC	*Rapporteur*
Bernard LAPEYRE	*Directeur de thèse*
Eric MOULINES	*Examinateur*
Gilles PAGES	*Rapporteur*
Denis TALAY	*Président du jury*
Bruno TUFFIN	*Examinateur*

Résumé

Cette thèse traite de deux sujets indépendants. La première partie est consacrée à l'étude des algorithmes stochastiques. Dans un premier chapitre introductif, je présente l'algorithme de Robbins et Monro [55] dans un parallèle avec l'algorithme de Newton pour l'optimisation déterministe. Ces quelques rappels permettent alors d'introduire les algorithmes stochastiques aléatoirement tronqués de Chen et Zhu [21] qui sont au cœur de cette thèse. La première étude de cet algorithme concerne sa convergence presque sûre qui est parfois établie sous des hypothèses assez changeantes. Ce premier chapitre est l'occasion de clarifier les hypothèses de la convergence presque sûre et d'en présenter une preuve simplifiée. Dans le second chapitre, nous poursuivons l'étude de cet algorithme en nous intéressant cette fois à sa vitesse de convergence. Plus exactement, nous considérons une version moyenne mobile de cet algorithme et démontrons un théorème centrale limite pour cette variante. Le troisième chapitre est consacré à deux applications de ces algorithmes à la finance : le premier exemple présente une méthode de calibration de la corrélation pour les modèles de marchés multidimensionnels alors que le second exemple poursuit les travaux de Arouna [7] en améliorant ses résultats.

La seconde partie de cette thèse s'intéresse à l'évaluation des options parisiennes en s'appuyant sur les travaux de Chesney, Jeanblanc-Picqué, et Yor [23]. La méthode d'évaluation se base sur l'obtention de formules fermées pour les transformées de Laplace des prix par rapport à la maturité. Nous établissons ces formules pour les options parisiennes simple et double barrières. Nous étudions ensuite une méthode d'inversion numérique de ces transformées. Nous établissons un résultat sur la précision de cette méthode numérique tout à fait performante. A cette occasion, nous démontrons également des résultats liés à la régularité des prix et l'existence d'une densité par rapport à la mesure de Lebesgues pour les temps parisiens.

mots clés : approximation stochastique, algorithmes tronqués, théorème centrale limite, options parisiennes, inversion numérique, transformées de Laplace.

Abstract

This thesis is split into two parts. The first one deals with the study of stochastic algorithms. In an introductory chapter, we present the Robbins and Monro [55] algorithm while making a parallel with the Newton algorithm commonly used in deterministic optimisation problems. These reminders naturally lead to the presentation of randomly truncated stochastic algorithms as first introduced by Chen and Zhu [21]. The first study of these randomly truncated stochastic algorithms is concerned with their almost sure convergence which has already been established under varying hypotheses. The first chapter gives us the opportunity to try to clarify the assumptions a little and to present a simplified proof of the almost sure convergence. The second chapter is devoted to the study of the convergence rate. More precisely, we consider a moving window version of the algorithm and establish a central limit theorem. The last chapter of this first part presents two applications of stochastic algorithms to finance. The first one deals with the calibration of the correlation in a multidimensional market model, while the second one is based on the work of Arouna [7]. Meanwhile, we improve the results Arouna had obtained.

The second part of the thesis is concerned with the pricing of Parisian options. The valuation technique is based on computing closed form formula for the Laplace transforms of the prices following the seminar work of Chesney, Jeanblanc-Picqué, et Yor [23] on the topic. First, we determine these formulae for the single barrier Parisian options following closely [23], second we do the same for double barrier Parisian options. Then, we study the numerical inversion of these Laplace transforms based on a contour integral technique. We establish the accuracy of the method we use. To do so, we prove the regularity of the Parisian option prices and establish the existence of a density w.r.t the Lebesgue measure for the “Parisian time”.

key words : stochastic approximation, truncated algorithms, central limit theorem, Parisian options, numerical inversion, Laplace transforms.

Remerciements

Je tiens tout d'abord à remercier Bernard Lapeyre mon directeur de thèse pour son soutien tout au long de ces trois années de thèse.

Je remercie Monique Jeanblanc et Gilles Pagès d'avoir accepté de rapporter sur cette thèse. J'ai particulièrement apprécié la lecture très minutieuse de Monique Jeanblanc et ses nombreuses remarques toujours constructives. Je suis également reconnaissant envers Gilles Pagès pour les différentes ouvertures qu'il a proposées.

J'exprime toute ma gratitude à Eric Moulines pour avoir accepté de faire partie du jury et pour les nombreuses discussions que nous avons eues sur les algorithmes stochastiques. Je remercie également Denis Talay et Bruno Tuffin qui ont eu a la gentillesse de prendre part au jury.

Naturellement, je tiens à saluer tous mes collègues du CERMICS et plus particulièrement ceux de l'équipe de "probabilités" Aurélien, Benjamin, Jean-François, Julien, Marian, Mohamed. A l'heure du changement, je pense à tout ceux qui ont partagé mon bureau durant ces trois ans Nicola, Pierre, Simone et Vincent. Je te tiens également à saluer Antonino pour tous les bons moments passés à travailler sur le logiciel PREMIA. Je voudrais ajouter une mention spéciale ç Aurélien pour son soutien et son aide scientifique si précieuse. Merci également à Jean-Philippe de m'avoir fait profité de sa grande expérience du développement logiciel.

De tout cœur, je remercie ma famille et surtout Céline pour leur soutien quotidien inestimable.

Table des matières

Introduction

La thèse que je présente ici se décompose en deux parties indépendantes que l'on pourrait rapprocher sous une seule et même étiquette : "méthodes numériques en probabilité". En effet dans chacune des deux parties, je m'intéresse à l'étude théorique de méthodes numériques. La première partie est consacrée à l'étude de méthodes d'approximation pour des problèmes d'optimisation dans un cadre stochastique. Quant à la seconde partie, elle s'articule autour du problème de l'évaluation des options parisiennes. Cette introduction a pour but de présenter le plus simplement possible une vue d'ensemble des différents résultats obtenus au cours de la thèse. Le lecteur soucieux du détail ne m'en voudra pas d'éluder volontairement les détails techniques qui sont par contre développés dans les chapitres respectifs.

Algorithmes stochastiques

Dans cette première partie, je me suis intéressé à l'étude théorique des algorithmes stochastiques. En première approximation, on pourrait dire que ces algorithmes sont des versions stochastiques des algorithmes déterministes comme l'algorithme de Newton ou l'algorithme du gradient. Nous verrons en fait plus loin que cette définition est un peu élémentaire. Nous intéressons au problème suivant : trouver le zéro $x^\star$ d'une fonction u définie sous forme d'espérance $u(x) = \mathbb{E}(U(x, Z))$. Lorsque l'on a une expression explicite de u, un algorithme déterministe comme l'algorithme de Newton permet d'approcher très efficacement $x^\star$ pour peu que u soit convexe. Malheureusement, il est assez courant que l'on ne dispose pas de formules fermées pour u. C'est dans ce cadre que les algorithmes stochastiques révèlent toute leur efficacité. On considère alors l'algorithme suivant

$$\underbrace{X_{n+1} = X_n - \gamma_{n+1} u(X_n)}_{\text{algorithme de Newton}} - \underbrace{\gamma_{n+1} \delta M_{n+1}}_{\text{terme de bruit}}. \tag{1}$$

La suite $(\gamma_n)_n$ est souvent appelée le pas ou le gain de l'algorithme. Le terme de bruit δM_n permet de prendre en compte le fait que l'on ne connaît pas la valeur de $u(X_n)$, on remplace alors cette valeur par une réalisation de $U(X_n, Z)$. Cet algorithme est connu depuis longtemps grâce à Robbins et Monro [55] et de nombreuses études de ses propriétés asymptotiques ont déjà été menées, un bref rappel en est fait au début du chapitre 1. Le comportement de cet algorithme se dégrade malheureusement de manière assez dramatique dès que la fonction u croit trop rapidement. De manière un peu schématique, on peut dire que cet algorithme perd ses propriétés de convergence dès lors que la fonction u a un comportement sur-linéaire, ce qui restreint considérablement son champ d'applications. C'est pourquoi nous nous intéressons ici à une version beaucoup plus robuste de cet algorithme présentée par Chen et Zhu [21]. Il s'agit tout simplement d'ajouter un terme de rejet (ou de projection entend-on parfois) à l'algorithme précédent pour se permettre de réinitialiser l'algorithme lorsque l'on estime qu'il varie de manière trop brutale d'un pas à l'autre.

$$\underbrace{X_{n+1} = X_n - \gamma_{n+1} u(X_n)}_{\text{algorithme de Newton}} - \underbrace{\gamma_{n+1} \delta M_{n+1}}_{\text{terme de bruit}} + \underbrace{\gamma_{n+1} p_{n+1}}_{\text{terme de rejet}} . \tag{2}$$

Dans la deuxième moitié du chapitre 1, nous présentons cet algorithme en détail et étudions les hypothèses minimales requises pour assurer sa convergence (voir la proposition 1.2.1). Cette discussion autour de la convergence de l'algorithme 2 est l'occasion de faire la synthèse des différents résultats existant dans la littérature. En particulier, la proposition 1.2.1 améliore le résultat de Chen et al. [22] et pour cette raison, nous avons jugé intéressant d'en donner une preuve complète et auto-suffisante.

Une fois la convergence établie, nous nous intéressons à la vitesse à laquelle elle a lieu : il s'agit d'étudier la convergence en loi de la suite $(X_n - x^\star)/\sqrt{\gamma_n}$. On parvient à démontrer que la limite suit une loi normale. Le résultat est certes déjà connu bien que sous des hypothèses différentes mais les démonstrations que l'on peut en trouver font toutes appel à des arguments de convergence fonctionnelle alors que la démonstration que nous en faisons ici s'appuie essentiellement sur le Théorème Central Limite pour les tableaux de martingales qui reste tout de même plus facile à appréhender (voir page 140 pour un énoncé). Nous réservons les arguments de type fonctionnel au chapitre 2 pour obtenir un résultat de convergence beaucoup plus fin.

La suite de l'étude est basée sur la remarque suivante : si une suite converge, alors sa moyenne de Césaro converge également et de manière beaucoup plus douce. Ce n'est pas tout à fait le concept de moyenne de Césaro que nous avons retenu mais plutôt celui de moyenne mobile, en effet la moyenne de Césaro converge certes de manière plus douce mais le régime asymptotique est atteint plus difficilement puisqu'il faut réussir à "oublier" les premiers termes de la somme. La moyenne mobile permet de conjuguer les avantages de la moyenne de Césaro (convergence plus douce) pourvu que la taille de la fenêtre soit bien choisie tout en ne subissant pas la pénalisation due aux premiers termes puisque par définition les premiers termes sont oubliés au fur et à mesure. On définit donc

$$\bar{X}_n = \frac{X_n + X_{n+1} + \cdots + X_{n+p_n}}{p_n}. \tag{3}$$

Etablir la convergence est immédiat. Le chapitre 2 est consacré à l'étude de sa vitesse de convergence. Nous verrons que si la suite p_n est bien choisie, alors la limite en loi lorsque n tend vers l'infini de $(\bar{X}_n - x^\star)/\sqrt{\gamma_n}$ existe et on trouve de nouveau une loi normale dont la variance est plus petite que pour l'algorithme non moyenné.

Cette partie sur les algorithmes stochastiques s'achève sur deux exemples d'application tirés de la finance. Le premier exemple montre comment utiliser un algorithme stochastique pour résoudre un problème de calibration. Dans le deuxième exemple introduit à l'origine par Arouna [7] nous expliquons comment utiliser les algorithmes stochastiques pour mettre en œuvre d'une méthode de fonction d'importance lors du calcul d'espérance dans un cadre gaussien.

Nous abordons tout d'abord un problème de calibration. La calibration est une véritable préoccupation pour tous les praticiens dont la résolution mathématique est souvent fort délicate. Le principe de la calibration est d'essayer étant donné un modèle mathématique de reproduire le plus fidèlement possible les prix des produits liquides fournis par le marché en ajustant au mieux les paramètres du modèle. Cet exercice amène à la résolution d'un problème de minimisation en grande dimension. Nous nous intéressons plus particulièrement au problème de la détermination de la corrélation dans les modèles de marché multidimensionnel. La formulation mathématique de ce problème entre parfaitement dans le cadre défini par les algorithmes stochastiques puisque la minimisation peut se ramener à trouver le (ou les) zéro(s) d'une fonction définie comme une espérance. Nous verrons d'ailleurs quelles hypothèses raisonnables assurent l'existence d'une unique solution à ce genre de problème.

Dans un second temps, nous nous intéressons à un problème de réduction de variance. Supposons que l'on veuille calculer par une méthode Monte Carlo $\mathbb{E}(f(X))$ et que l'on dispose d'une famille de variables aléatoires $\mathcal{X} = \{Y_\theta \ : \ \theta \in \mathbb{R}\}$ telles que $\mathbb{E}(Y_\theta) = \mathbb{E}(f(X))$ pour tout θ. En vertu du Théorème Central Limite, il faut bien sûr utiliser la variable aléatoire de plus petite variance dans la méthode de Monte Carlo. On cherche donc dans un premier temps $\theta^\star$ telle que $\mathrm{Var}(Y_{\theta^\star}) = \min_\theta \mathrm{Var}(Y_\theta)$. Puis on approche $\mathbb{E}(f(X))$ par $\frac{1}{n}\sum_{i=1}^n Y_{\theta^\star}^{(i)}$. Pour mettre en œuvre une telle méthode, il faut donc être capable de construire une telle famille $\mathcal{X}$. Dans un cadre gaussien, il suffit par exemple de considérer $Y_\theta = f(X+\theta)\, \mathrm{e}^{-\theta\cdot X - \frac{\theta^2}{2}}$. $\theta^\star$ s'écrit alors comme l'unique zéro d'une espérance. C'est précisément à cet instant que les algorithmes stochastiques entrent en jeu. Ils permettent d'implémenter une méthode d'approximation simple et efficace du paramètre optimal $\theta^\star$. Nous avons expérimenté cette méthode dans le cadre du pricing de produits dérivés sur actions et les résultats obtenus parlent d'eux-mêmes. Pour les options considérées, la variance de la méthode de Monte Carlo utilisant le paramètre optimal est réduite de manière significative. La méthode développée présente également l'avantage non négligeable de pouvoir être mise en œuvre de manière presque totalement automatique, ce qui est bien sûr un atout fort important pour les praticiens. Les différentes méthodes décrites dans ce chapitre 3 ont été implémentées en C++ à l'aide de la librairie de calcul matriciel *Blas*.

Options Parisiennes

La deuxième partie de cette thèse s'inscrit plus encore que la première dans le cadre des mathématiques financières. Nous nous sommes intéressés à l'évaluation d'options barrières un peu particulières : les options parisiennes. A la différence des options barrières qui sont activées (ou désactivées) si le cours du sous-jacent franchit un niveau fixé à l'avance, les options parisiennes s'intéressent au temps que le sous-jacent passe de manière continue au dessus (ou dessous) d'un niveau donné. De part la nature même des ces options, leur évaluation est fort complexe et aucune formule fermée ne peut être obtenue. De nombreuses méthodes d'approximation existent mais faut-il encore qu'elles soient suffisamment rapides et précises pour être utilisées pour calculer la couverture de telles options, ce qui reste quand même le but ultime de tout praticien.

Les chapitres 4 et 5 sont consacrés à l'approximation du prix des options parisiennes. Le chapitre 4 traite des options parisiennes à simple barrière tandis que le chapitre 5 se concentre sur les options parisiennes à double barrière : pour ces options le payoff n'est activé (ou au contraire désactivé) que si la valeur du sous-jacent ne sort d'un intervalle donné plus d'une certaine durée. La méthode étudiée pour approcher le prix de ces options est de calculer les transformées de Laplace par rapport à la maturité des différents prix pour lesquels nous obtenons des formules fermées. La complexité des formules obtenues pour ces transformées de Laplace interdit tout espoir de trouver des formules fermées pour les prix eux-mêmes. Il faut alors se rabattre sur une inversion numérique des transformées de Laplace des prix. Le principe est de proposer un algorithme permettant d'évaluer numériquement le prix pour la maturité qui nous intéresse. Notre choix pour réaliser cette inversion numérique s'est portée sur la méthode proposée par Abate et al. [2]. Il s'agit en fait de discrétiser une intégrale dans le plan complexe. Pour le problème qui nous intéresse, nous sommes parvenus à établir un résultat sur la précision de la méthode.

Dans cette partie, nous donnons les différentes formules de transformées de Laplace des prix ainsi que l'ensemble des relations de parité permettant de déduire les autres transformées. Ainsi, avec les formules présentées dans cette thèse, il est tout à fait possible de calculer numériquement les prix de toutes les options parisiennes qu'elles soient à simple ou double barrière. Pour les options parisiennes simple barrière, nous expliquons également comment calculer les prix à tout instant à partir de la date d'émission de l'option, ce qui est indispensable pour pouvoir calculer la couverture de l'option.

L'intégralité des formules présentées dans cette partie ainsi que l'algorithme d'inversion des transformées de Laplace ont été implémentées en C dans le logiciel PREMIA [51].

Notations

In this manuscript, we use the following standard notations. The prime notation (symbol $'$) stands for the transpose operator, we write A' or $(A)'$ interchangeably. For any element of $\mathbb{R}^d$, the notation $\|\cdot\|$ denotes the standard Euclidean norm and $(\cdot|\cdot)$ the associated scalar product. If needed, we would also use the notation $\|\cdot\|$ for the algebra norm on the square matrices defined by $\|A\| = \sup_{\|x\|=1} \|Ax\|$. The elements of $\mathbb{R}^d$ are always supposed to be encoded as column vectors. We may also write $u \cdot v$ for the scalar product of the vectors v and u.

If A is a matrix, $A(:\,, i)$ (resp. $A(j, :\,)$) denotes the $i-$th column (resp. $j-$th row) of A.

Première partie

Algorithmes Stochastiques

Chapitre 1

Les algorithmes stochastiques en bref

Ce chapitre a pour but de présenter les algorithmes stochastiques dans leur ensemble en prenant comme point de départ leurs pendants déterministes. Dans un premier temps, nous rappelons une version élémentaire de l'algorithme de Newton dont nous dérivons ensuite une version perturbée qui est la base même de l'approximation stochastique. Nous présentons ensuite logiquement l'algorithme de Robbins Monro. Une fois ces bases posées, nous nous arrêtons sur le problème de l'optimisation stochastique sous contraintes et rappelons pour les algorithmes stochastiques sous contraintes un résultat de convergence trajectorielle puis un résultat sur leur vitesse de convergence.

La dernière partie de ce chapitre est consacrée aux cas des fonctions à croissance rapide qui prennent en défaut l'algorithme de Robbins Monro. Il faut alors recourir à des raffinements comme la troncature à des niveaux aléatoires proposée par Chen et Zhu [21]. Pour ce dernier algorithme, nous démontrons un résultat de convergence presque sûre ainsi qu'un théorème central limite. Ces résultats sont certes déjà connus sous des hypothèses plus restrictives et il nous parait important de comprendre en profondeur ces deux résultats avant d'attaquer les chapitres suivants.

1.1 Les algorithmes stochastiques

1.1.1 un détour déterministe

Considérons le problème de recherche de zéros d'une fonction u donnée dans un cadre déterministe. On peut par exemple utiliser le résultat suivant adapté de Duflo [29, Proposition 1.2.3]

Proposition 1.1.1. *Soit u une fonction continue à valeurs réelles telle que*
- *il existe un unique $x^\star$ tel que $u(x^\star) = 0$,*
- *pour tout $x \neq x^\star$, $u(x)(x - x^\star) > 0$,*
- *il existe $K > 0$, tel que $|u(x)| \leq K(1 + |x|)$.*

Soit $(\gamma_n)_n$ *une suite positive décroissante tendant vers* 0 *telle que la série* $\sum \gamma_n$ *diverge. Alors, la suite* $(x_n)_n$ *définie par*

$$x_{n+1} = x_n - \gamma_n u(x_n)$$

converge vers $x^\star$ *pour toutes valeurs initiales* x_0.

On peut également citer d'autres algorithmes beaucoup plus évolués tels que l'algorithme du gradient conjugué ou encore l'algorithme du simplexe. Néanmoins, l'algorithme décrit par la proposition 1.1.1 est très facile à mettre en œuvre et sert de point de départ aux algorithmes utilisés dans un contexte aléatoire. Pour une description des différents algorithmes couramment utilisés en optimisation déterministe, le lecteur pourra se référer au livre de Bonnans et al. [16].

Dans de nombreuses situations la fonction dont on cherche le zéro n'est connue qu'à une perturbation près. La recherche des zéros par les méthodes d'optimisation déterministe devient alors plus périlleuse et l'on a recours à des algorithmes stochastiques comme celui introduit par Robbins et Monro [55].

1.1.2 L'algorithme de Robbins Monro

On cherche toujours à trouver l'ensemble

$$u^{-1}(0) = \{x \in \mathbb{R}^d \ : \ u(x) = 0\} \tag{1.1}$$

des zéros d'une fonction u définie de $\mathbb{R}^d$ dans $\mathbb{R}^d$. On suppose ne connaître la fonction u qu'en “moyenne” ou qu'à une perturbation près, c'est-à-dire que pour x donné dans $\mathbb{R}^d$, $u(x)$ lui-même n'est pas observable mais seulement une quantité $Y(x) = u(x) + \varepsilon(x)$ où $(\varepsilon(x), x \in \mathbb{R}^d)$ est une famille de variables aléatoires centrées i.i.d à valeurs dans $\mathbb{R}^d$. Ainsi $\mathbb{E}(Y(x)) = u(x)$. Dans une telle situation, les algorithmes mis en œuvre s'écrivent

$$X_{n+1} = X_n - \gamma_{n+1} Y_{n+1}, \tag{1.2}$$

où $Y_{n+1} = (u(X_n) + \varepsilon_{n+1})$ avec $(\varepsilon_n)_n$ une suite de variables aléatoires i.i.d telle que ε_{n+1} soit indépendante de $\mathcal{F}_n = \sigma(X_0, \ldots, X_n)$. La suite $(\gamma_n)_n$ est strictement positive et décroissante.

Sous certaines conditions, il existe des résultats sur la convergence presque sûre de la suite $(X_n)_n$ vers les zéros de u. Dans le cas où l'ensemble $u^{-1}(0)$ n'est pas réduit à un singleton, on peut citer les résultats de Pelletier [49], Delyon [27] ou Benaïm [11, 12]. Lorsque u admet un unique zéro, on peut par exemple consulter les travaux de Robbins et Monro [55] ou Duflo [29] pour trouver une démonstration du théorème suivant. Ce résultat a aussi été démontré en utilisant la technique des équations différentielles ordinaires (ODE) par Borkar et Meyn [17] par exemple. Cette technique et également utilisée par Fort et Pagès [31] qui étudient le lien entre les trajectoires de l'ODE et la convergence de l'algorithme associé.

Théorème 1.1.2. *Soit $(Y_n)_n$ une suite de variables aléatoires de carré intégrable et $(X_n)_n$ la suite définie par (1.2). Supposons qu'il existe une fonction u continue satisfaisant*

$$u(X_n) = \mathbb{E}(Y_{n+1}|\mathcal{F}_n),$$

et telle que

(A1.1) *i. il existe un unique $x^\star$ tel que $u(x^\star) = 0$ et pour tout $x \neq x^\star$, $(u(x)|(x - x^\star)) > 0$,*

ii. $\mathbb{E}(\|Y_{n+1}\|^2 |\mathcal{F}_n) \leq K(1 + \|X_n\|^2)$ p.s.

(A1.2) *$\sum_n \gamma_n = \infty$ et $\sum_n \gamma_n^2 < \infty$.*

Alors, $X_n \xrightarrow{p.s.} x^\star$.

Quelques remarques sur les hypothèses. L'hypothèse ($\mathcal{A}$1.1-*i)* est une hypothèse de type Lyapounov qui assure l'unicité du zéro de u et qui le plus souvent revient à dire que la fonction u dérive d'une fonction strictement convexe, ce qui est un cadre tout à fait raisonnable pour envisager un problème d'optimisation. La seconde hypothèse s'exprime comme une contrainte sur la manière de choisir la suite de pas de l'algorithme mais ne dépend nullement de la fonction u. Les véritables difficultés apparaissent avec l'hypothèse ($\mathcal{A}$1.1-*ii)* qui en quelque sorte impose qu'en "moyenne" la fonction u ait un comportement sous linéaire. En pratique, cette hypothèse est fort contraignante et rarement satisfaite. Pour palier ce problème, des versions modifiées de cet algorithme ont été proposées par Chen et Zhu [21] comme nous le verrons au paragraphe 1.1.4.

Le cadre général décrit précédemment s'applique également lorsque la fonction u est donnée sous la forme d'une espérance

$$u(x) = \mathbb{E}(U(x, Z)), \tag{1.3}$$

où Z est une variable aléatoire à valeurs dans $\mathbb{R}^n$. Dans ce cas, il suffit de considérer une suite $(Z_n)_n$ i.i.d. selon la loi de Z et de poser $Y_{n+1} = U(X_n, Z_{n+1})$ avec Z_{n+1} indépendante de la tribu $\mathcal{F}_n$. On vérifie facilement que $\mathbb{E}(U(X_n, Z_{n+1})|\mathcal{F}_n) = u(X_n)$. On est donc encore parfaitement dans le cadre du théorème 1.1.2. Puisque les deux approches se présentent finalement de la même manière, on supposera désormais, dans un souci de cohérence, que u est définie par l'équation (1.3).

Il arrive fréquemment que le problème (1.1) soit posé sous contraintes, ce qui correspond au cas de l'optimisation sous contraintes puisqu'il ne faut pas oublier que la résolution de (1.1) est généralement motivée par un problème d'optimisation.

1.1.3 Les algorithmes sous contraintes

Cette fois, on cherche à déterminer l'ensemble

$$u^{-1}(0) = \{x \in \mathcal{H} \;:\; u(x) = 0\} \tag{1.4}$$

où $\mathcal{H}$ est un sous ensemble convexe fermé de $\mathbb{R}^n$ qui représente l'ensemble des contraintes. Si l'ensemble $u^{-1}(0)$ est réduit à un unique élément, on peut être tenté d'utiliser l'algorithme précédent. Comment peut-on alors assurer que la suite X_n converge vers un élément de $\mathcal{H}$? Une solution consiste par exemple à projeter à chaque itération le vecteur X_n sur l'ensemble $\mathcal{H}$. Ainsi la nouvelle suite obtenue, si elle converge, converge vers un élément de $\mathcal{H}$ puisque $\mathcal{H}$ est fermé. Reste à savoir si cette nouvelle suite converge et si oui, converge-t-elle vers l'unique $x^\star$ satisfaisant $u(x^\star) = 0$.

Pour $X_0 \in \mathcal{H}$, on considère donc la suite $(X_n)_n$ définie par

$$X_{n+1} = \Pi\left(X_n - \gamma_{n+1} U(X_n, Z_{n+1})\right), \tag{1.5}$$

où Π est la projection Euclidienne sur $\mathcal{H}$. Ce nouvel algorithme sous contraintes a été étudié par Dupuis et Kushner [30] et Buche et Kushner [19] entre autres.
L'équation (1.5) peut alors se réécrire

$$X_{n+1} = X_n - \gamma_{n+1} u(X_n) - \gamma_{n+1}\delta M_{n+1} + \gamma_{n+1} p_{n+1}, \tag{1.6}$$

où

$$\begin{aligned} \delta M_{n+1} &= U(X_n, Z_{n+1}) - u(X_n), \\ \gamma_{n+1} p_{n+1} &= \Pi(X_n - \gamma_{n+1} U(X_n, Z_{n+1})) - (X_n - \gamma_{n+1} U(X_n, Z_{n+1})). \end{aligned}$$

Convergence presque sûre

La convergence presque sûre de ces algorithmes sous contraintes s'obtient de manière assez similaire à celle de l'algorithme de Robbins Monro et est clairement exposée dans les livres de Kushner et Clark [42] et Kushner et Yin [44]. On peut énoncer le théorème suivant

Théorème 1.1.3. *Soit $(X_n)_n$ la suite définie par (1.6). Supposons que l'hypothèse ($\mathcal{A}$1.1-i) soit vérifiée sur $\mathcal{H}$ et que la suite $(\gamma_n)_n$ satisfasse ($\mathcal{A}$1.2). Si de plus*

($\mathcal{A}$1.3) $\mathbb{E}(\|U(x,Z)\|^2) \leq K(1 + \|x\|^2)$ *pour tout x dans $\mathcal{H}$,*
alors $X_n \xrightarrow{p.s.} x^\star$.

Remarque 1.1.4. *Dans l'équation (1.5), on pourrait en fait se contenter de réinitialiser l'algorithme à n'importe quel point de $\mathcal{H}$ de manière déterministe, voire même mesurable par rapport à $\mathcal{F}_n$ et le théorème 1.1.3 resterait valable. Cette latitude est très appréciable dans les applications. On peut par exemple, décider d'implémenter un algorithme d'acceptation rejet en posant $X_{n+1} = X_n$ si $X_n - \gamma_{n+1} U(X_n, Z_{n+1}) \notin \mathcal{H}$.*

La convergence des algorithmes sous contraintes a été étudiée plus en détails par Buche et Kushner [19] qui se sont intéressés à la vitesse de convergence de X_n vers $x^\star$ en établissant un résultat de type théorème central limite.

Vitesse de convergence

La vitesse de convergence des algorithmes stochastiques, qu'ils soient sous contraintes ou non, s'étudie en considérant la suite des erreurs renormalisées

$$\Delta_n = \frac{X_n - x^\star}{\sqrt{\gamma_n}}. \tag{1.7}$$

Il s'agit alors d'établir un résultat de convergence en loi pour la suite $(\Delta_n)_n$ à la manière du théorème central limite. Il existe autour de la vitesse de convergence des algorithmes stochastiques une littérature très abondante tant les manières d'aborder le problème sont nombreuses. Pour une approche plus algébrique, on consultera Delyon [27, 28], alors que le lecteur plus familier des arguments relatifs à la convergence fonctionnelle préférera se tourner vers les travaux de Buche et Kushner [19]. Ces approches plus fonctionnelles avaient déjà été développées pour l'algorithme de Robbins Monro par Bouton [18] puis par Benveniste et al. [13]. L'algorithme de Robbins Monro est parfois étudié avec un pas constant (i.e. $\gamma_n = \gamma$) et l'on s'intéresse au comprtement lorsque γ tend vers 0. Pour l'étude du comportement asymptotique de tels algorithmes à pas constant l'algorithme, on renvoie à Fort et Pagès [32] par exemple.

Nous allons maintenant énoncer les deux résultats relatifs à la vitesse de convergence des algorithmes stochastiques sous contraintes dans le cas particulier où la suite $(\gamma_n)_n$ s'écrit

$$\gamma_n = \frac{\gamma}{n^\alpha} \tag{1.8}$$

avec $1/2 < \alpha \leq 1$. Ces deux bornes sur l'exposant α permettent de s'assurer que la suite $(\gamma_n)_n$ vérifie l'hypothèse ($\mathcal{A}1.2$). La valeur $\alpha = 1$ doit être considérée séparément puisqu'elle conduit à une limite différente comme nous allons le voir dans la suite.

Introduisons les jeux d'hypothèses suivants

($\mathcal{A}$1.4) *i.* $\forall x \in \mathcal{H},\ x \neq x^\star,\ (x - x^\star | u(x)) > 0.$

ii. Il existe une fonction $y : \mathbb{R}^d \to \mathbb{R}^{d\times d}$ vérifiant $\lim_{\|x\|\to 0} \|y(x)\| = 0$ et une matrice A symétrique définie positive telle que

$$u(x) = A(x - x^\star) + y(x - x^\star)(x - x^\star).$$

($\mathcal{A}$1.5) *i.* Il existe un réel $\rho > 0$ tel que

$$\kappa = \sup_n \mathbb{E}\left(\|\delta M_n\|^{2+\rho}\right) < \infty.$$

nous posons $\kappa_0 = \sup_n \mathbb{E}\left(\|\delta M_n\|^2\right)$.

ii. Il existe une matrice Σ symétrique définie positive telle que

$$\mathbb{E}\left(\delta M_n \delta M_n' | \mathcal{F}_{n-1}\right) \xrightarrow[n\to\infty]{\mathbb{P}} \Sigma.$$

($\mathcal{A}$1.6) $x^\star$ appartient à l'intérieur de $\mathcal{H}$.

Un TCL pour $1/2 < \alpha < 1$

Théorème 1.1.5. *Sous les hypothèses ($\mathcal{A}$1.3), ($\mathcal{A}$1.4), ($\mathcal{A}$1.5) et ($\mathcal{A}$1.6), la suite* $(\Delta_n)_n$ *converge en loi vers une variable aléatoire normale centrée et de matrice de covariance*

$$V = \int_0^\infty \exp(-At)\Sigma \exp(-At)dt.$$

Un TCL pour $\alpha = 1$

Théorème 1.1.6. *Sous les hypothèses ($\mathcal{A}$1.3), ($\mathcal{A}$1.4), ($\mathcal{A}$1.5) et ($\mathcal{A}$1.6) et si de plus* $\gamma A - \frac{I}{2}$ *est définie positive, la suite* $(\Delta_n)_n$ *converge en loi vers une variable aléatoire normale centrée et de matrice de covariance*

$$V = \gamma \int_0^\infty \exp\left(\left(\frac{I}{2} - \gamma A\right) t\right)\Sigma \exp\left(\left(\frac{I}{2} - \gamma A\right) t\right)dt.$$

On peut trouver une démonstration de ces deux résultats dans les travaux de Delyon [27] ou de Buche et Kushner [19] par exemple. Pour une méthode permettant d'estimer la matrice de covariance asymptotique, on pourra se référer aux travaux de Glynn et Hsieh [35].

Dans le cas non contraint, l'hypothèse ($\mathcal{A}$1.1-*ii)* est une condition très forte qu'il est souvent bien difficile de satisfaire en pratique. Le lecteur pourra consulter le chapitre 3 pour découvrir un exemple d'utilisation des algorithmes stochastiques dans un contexte où la fonction U ne vérifie pas l'hypothèse ($\mathcal{A}$1.1-*ii)* de croissance sous linéaire en "moyenne".

1.1.4 L'amélioration proposée par Chen

Nous considérons toujours le problème (1.1) où la fonction u est définie par l'équation (1.3). Supposons de plus que la condition ($\mathcal{A}$1.1-*ii)* ne soit pas satisfaite, c'est-à-dire que $\mathbb{E}(\|U(x,Z)\|^2)$ croisse plus vite que $\|x\|^2$. Comme on vient de le voir précédemment, l'algorithme de Robbins Monro standard est pris en défaut dans une telle situation, il faut alors recourir à des procédures plus élaborées mais aussi plus délicates à mettre en œuvre et à appréhender d'un point de vue mathématique.

L'algorithme que nous allons étudier ici a été introduit par Chen et Zhu [21]. Le principe est de modifier l'algorithme de Robbins Monro (équation (1.2)) de manière à éviter que la suite $(X_n)_n$ n'explose pendant les premières itérations.

Considérons une suite croissante de compacts $(\mathcal{K}_j)_j$ tels que

$$\bigcup_{j=0}^{\infty} \mathcal{K}_j = \mathbb{R}^d \quad \text{et} \quad \forall j,\ \mathcal{K}_j \subsetneq \mathrm{int}(\mathcal{K}_{j+1})\,. \tag{1.9}$$

De nouveau, $(Z_n)_n$ est une suite de variables aléatoires i.i.d. selon la loi de Z et $(\gamma_n)_n$ une suite positive et décroissante. Pour $X_0 \in \mathcal{K}_0$ et $\sigma_0 = 0$, nous définissons les suites de variables aléatoires $(X_n)_n$ et $(\sigma_n)_n$.

$$\begin{cases} & X_{n+\frac{1}{2}} = X_n - \gamma_{n+1} U(X_n, Z_{n+1}), \\ \text{si } X_{n+\frac{1}{2}} \in \mathcal{K}_{\sigma_n} & X_{n+1} = X_{n+\frac{1}{2}} \quad \text{and} \quad \sigma_{n+1} = \sigma_n, \\ \text{si } X_{n+\frac{1}{2}} \notin \mathcal{K}_{\sigma_n} & X_{n+1} = X_0 \quad \text{and} \quad \sigma_{n+1} = \sigma_n + 1. \end{cases} \tag{1.10}$$

Remarque 1.1.7. *Quand $X_{n+\frac{1}{2}} \notin \mathcal{K}_{\sigma_n}$, on peut en fait réinitialiser X_{n+1} avec n'importe quelle fonction mesurable de $(X_0, \ldots, X_n)$ à valeurs dans un compact fixé (indépendant de n). L'existence d'un tel compact est tout à fait essentielle pour démontrer la convergence presque sûre de $(X_n)_n$.*

Nous définissons $\mathcal{F}_n = \sigma(Z_k; k \leq n)$ la tribu engendrée par les vecteurs aléatoires $(Z_k, k \leq n)$. Notons que X_n est $\mathcal{F}_n-$mesurable puisque X_0 est déterministe et U est supposée mesurable.

Il est souvent plus commode, comme dans le cas des algorithmes sous contraintes, de réécrire (1.10) comme suit

$$X_{n+1} = X_n - \gamma_{n+1} u(X_n) - \gamma_{n+1} \delta M_{n+1} + \gamma_{n+1} p_{n+1} \tag{1.11}$$

où

$$\begin{aligned} \delta M_{n+1} &= U(X_n, Z_{n+1}) - u(X_n), \\ \text{et} \quad p_{n+1} &= \begin{cases} u(X_n) + \delta M_{n+1} + \frac{1}{\gamma_{n+1}}(X_0 - X_n) & \text{si } X_{n+\frac{1}{2}} \notin \mathcal{K}_{\sigma_n}, \\ 0 & \text{sinon.} \end{cases} \end{aligned}$$

1.2 Résultats de convergence pour l'algorithme de Chen

1.2.1 Convergence presque sûre

De nombreux résultats relatifs à la convergence presque sûre de l'algorithme tronqué de Chen existent sous des hypothèses parfois assez différentes. On citera en particulier les travaux de Chen et Zhu [21], Delyon [28] et Andrieu et al. [5].

Nous présentons une amélioration du résultat de Chen et Zhu [21] qui démontrent la convergence presque sûre sous une hypothèse de convergence globale de la série $\sum_n \gamma_{n+1} \delta M_{n+1}$ alors que nous nous contentons d'une condition de convergence locale nettement plus facile à vérifier en pratique.

Proposition 1.2.1. *Sous les hypothèses ($\mathcal{A}$1.2), ($\mathcal{A}$1.4-i) et si*

($\mathcal{A}$1.7) *pour tout $q > 0$, la série $\sum_n \gamma_{n+1} \delta M_{n+1} \mathbf{1}_{\{\|X_n - x^\star\| \leq q\}}$ converge presque sûrement,*

alors la suite $(X_n)_n$ converge p.s. vers $x^\star$ et de plus la suite $(\sigma_n)_n$ est finie p.s. (i.e. pour n assez grand $p_n = 0$ p.s.).

Corollaire 1.2.2. *Sous les hypothèses ($\mathcal{A}$1.2), ($\mathcal{A}$1.4-i) et si la fonction* $x \longmapsto \mathbb{E}(\|U(x,Z)\|^2)$ *est bornée sur tout compact, alors la suite* $(X_n)_n$ *converge p.s. vers* $x^\star$ *et de plus la suite* $(\sigma_n)_n$ *est finie p.s. (i.e. pour* n *assez grand* $p_n = 0$ *p.s.).*

Démonstration. Il suffit de vérifier que l'hypothèse ($\mathcal{A}$1.7) de la proposition 1.2.1 est vérifiée. Pour cela, on utilise le théorème de convergence de martingales de carré intégrable. En effet si on pose $M_n = \sum_{i=1}^n \gamma_i \delta M_i \mathbf{1}_{\{\|X_{i-1}-x^\star\| \le q\}}$, $(M_n)_n$ est bien une martingale. De plus, $\langle M \rangle_n = \sum_{i=1}^n \gamma_i^2 \mathbb{E}(\delta M_i \delta M_i' | \mathcal{F}_{i-1}) \mathbf{1}_{\{\|X_{i-1}-x^\star\| \le q\}}$. Ainsi puisque $\sum_i \gamma_i^2 < \infty$, la série $\sum_n \gamma_{n+1} \delta M_{n+1} \mathbf{1}_{\{\|X_n - x^\star\| \le q\}}$ converge presque sûrement. ■

La démonstration de la proposition 1.2.1 repose sur le lemme suivant qui donne une condition pour que la suite $(X_n)_n$ soit presque sûrement dans un compact.

Lemme 1.2.3. *Si pour tout* $q > 0$*, la série* $\sum_{n>0} \gamma_n \delta M_n \mathbf{1}_{\{\|X_{n-1}-x^\star\| < q\}}$ *converge presque sûrement et si* $p_n \mathbf{1}_{\{\|X_{n-1}-x^\star\| < q\}} \longrightarrow 0$*, alors la suite* $(X_n)_n$ *reste presque sûrement dans un compact.*

Dans le lemme 1.2.3, le compact auquel il est fait allusion est aléatoire, en particulier ce lemme n'implique pas que le nombre de projections soit borné indépendemment de l'aléa ω.

Démonstration de la proposition 1.2.1. La démonstration se déroule en deux étapes.

- Soit $q > 0$.
  On définit $\bar{M}_n = \sum_{i=1}^n \gamma_i \delta M_i \mathbf{1}_{\{\|X_{i-1}-x^\star\| \le q\}}$. $\bar{M}_n$ converge presque sûrement grâce à l'hypothèse ($\mathcal{A}$1.7).
  Supposons que $\sigma_n \longrightarrow \infty$. Ceci est en contradiction avec la conclusion du lemme 1.2.3, ce qui implique que l'hypothèse $p_n \mathbf{1}_{\{\|X_{n-1}-x^\star\| < q\}}$ tend vers 0 n'est pas vérifiée.

  $$\exists\, \eta > 0,\ q > 0, \quad \forall N > 0,\ \exists n > N \quad \mathbf{1}_{\{\|X_n - x^\star\| \le q\}} \left\|p_{n+1}\right\| > \eta.$$

  Soit $\varepsilon > 0$. Il existe donc une sous-suite $X_{\phi(n)}$ telle que pour tout $n > 0$, $\mathbf{1}_{\{\|X_{\phi(n)}-x^\star\| \le q\}} \left\|p_{\phi(n)+1}\right\| \neq 0$ et $\left\|\gamma_{\phi(n)+1} \delta M_{\phi(n)+1}\right\| \le \varepsilon$.
  Ainsi donc $\left\|X_{\phi(n)} - x^\star\right\| \le q$ et pourtant le nouvel itéré potentiel $X_{\phi(n)+\frac{1}{2}} = X_{\phi(n)} - \gamma_{\phi(n)+1}(u(X_{\phi(n)}) + \delta M_{\phi(n)+1})$ n'appartient pas à $\mathcal{K}_{\sigma_{\phi(n)}}$. Comme la fonction u est continue, la quantité $\left\|\gamma_{\phi(n)+1} u(X_{\phi(n)})\right\|$ peut être rendue plus petite que ε. Par hypothèse $\left\|\gamma_{\phi(n)+1} \delta M_{\phi(n)+1}\right\| \le \varepsilon$ et comme $\left\|\gamma_{\phi(n)+1} u(X_{\phi(n)})\right\| \le \varepsilon$, il suffit de choisir $\varepsilon < 1/2$ pour assurer que

  $$\left\|X_{\phi(n)} - x^\star - \gamma_{\phi(n)+1}(u(X_{\phi(n)}) + \delta M_{\phi(n)+1})\right\| \le q + 1.$$

  Soit l le plus petit entier tel que $B(x^\star, q+1) \subset \mathcal{K}_l$ (il existe d'après (1.9)), alors $\sigma_{\phi(n)} < l$ pour tout n. Puisque la suite $(\sigma_n)_n$ est croissante, ceci termine la première étape de la démonstration qui montre que $\limsup_n \sigma_n < \infty$ p.s..

- D'après le point précédent $\limsup_n \sigma_n < \infty$ p.s.. Ainsi, la suite $(X_n)_n$ est presque sûrement compacte. On peut a posteriori prendre $q = \infty$ dans l'hypothèse ($\mathcal{A}1.7$) et dire que $\sum_i \gamma_i \delta M_i$ converge presque sûrement. Considérons désormais

$$X'_n = X_n - \sum_{i=n+1}^{\infty} \gamma_i \delta M_i.$$

Puisque la série $\sum_{i>0} \gamma_i \delta M_i$ converge p.s. et que X_n reste dans un compact, X'_n reste également dans un compact. Notons $\mathcal{C}$ ce compact. Posons $\bar{u} = \sup_{x\in\mathcal{C}} \|u(x)\|$.

$$X'_{n+1} = X'_n - \gamma_{n+1} u(X'_n) + \gamma_{n+1}\varepsilon_n,$$

où $\varepsilon_n = u(X'_n) - u(X_n)$. Comme $\|X'_n - X_n\| \longrightarrow 0$ et que u est continue, $\|\varepsilon_n\| \longrightarrow 0$.

$$\begin{aligned}\left\|X'_{n+1} - x^\star\right\|^2 \leq \|X'_n - x^\star\|^2 & -2\gamma_{n+1}(X'_n - x^\star \mid u(X'_n)) \\ & + \gamma_{n+1}^2(\varepsilon_n^2 + \bar{u}^2) - 2\gamma_{n+1}(X'_n - x^\star \mid \varepsilon_n).\end{aligned}$$

Nous pouvons réécrire cette inégalité en introduisant une suite $\varepsilon'_n \longrightarrow 0$.

$$\left\|X'_{n+1} - x^\star\right\|^2 \quad \leq \quad \|X'_n - x^\star\|^2 -2\gamma_{n+1}(X'_n - x^\star \mid u(X'_n)) + \gamma_{n+1}\varepsilon'_n. \quad (1.12)$$

Soit $\delta > 0$. Si $\|X'_n - x^\star\|^2 > \delta$, alors $(X'_n - x^\star\ ,\ u(X'_n)) > c > 0$. Par conséquent, pour n assez grand l'équation (1.12) devient

$$\left\|X'_{n+1} - x^\star\right\|^2 \quad \leq \quad \|X'_n - x^\star\|^2 - \gamma_{n+1} c \mathbf{1}_{\{\|X'_n - x^\star\|^2 > \delta\}} + \gamma_{n+1}(\bar{c} + \varepsilon'_n)\mathbf{1}_{\{\|X'_n - x^\star\|^2 \leq \delta\}},$$

où $\bar{c} = \sup_{\|x-x^\star\|^2\leq\delta}(x - x^\star | u(x))$. Comme $\sum_n \gamma_n = \infty$, chaque fois que $\|X'_n - x^\star\|^2 > \delta$, la suite X'_n est ramenée dans la boule $\bar{B}(x^\star, \sqrt{\delta})$ en un nombre fini d'itérations. Ainsi, pour n assez grand

$$\|X'_n - x^\star\|^2 < \delta + \gamma_{\phi(n)+1}(\bar{c} + \varepsilon'_{\phi(n)}),$$

où $\phi(n) = \sup\{p \leq n; \left\|X'_p - x^\star\right\|^2 \leq \delta\}$. Comme $\phi(n)$ tend p.s. vers l'infini avec n, $\limsup_n \|X'_n - x^\star\|^2 \leq \delta$ pour tout $\delta > 0$. Ceci prouve que $X'_n \longrightarrow x^\star$. Puisque la série $\sum_n \gamma_{n+1}\delta M_{n+1}$ converge, nous avons également $X_n \longrightarrow x^\star$.

■

Nous allons à présent démontrer le lemme 1.2.3.

Démonstration du lemme 1.2.3. Si $\sigma_n < \infty$ p.s., alors la conclusion du lemme est évidente. Supposons donc que $\sigma_n \longrightarrow \infty$. Comme chaque fois que σ_n augmente, la suite X_n est réinitialisée à un point fixe de $\mathcal{K}_0$, l'existence d'un compact dans lequel la suite revient infiniment souvent est immédiate.

Soit $M > 0$ tel que la suite $(X_n)_n$ revienne infiniment souvent dans l'ensemble $\mathcal{C}$ définit par $\mathcal{C} = \{x : \|x - x^\star\|^2 \leq M\}$. Nous pouvons réécrire les hypothèses du lemme comme suit

$$\forall \varepsilon > 0, \exists N > 0 \text{ t.q. } \forall n, p \geq N \text{ on a } \begin{cases} \left\| \sum_{k=n}^{p} \gamma_k \delta M_k \mathbf{1}_{\{\|X_{k-1} - x^\star\|^2 \leq M+2\}} \right\| < \varepsilon, \\ \gamma_n < \varepsilon, \\ \mathbf{1}_{\{\|X_{n-1} - x^\star\|^2 \leq M+2\}} \|p_n\| < \varepsilon. \end{cases} \tag{1.13}$$

Soit $\varepsilon > 0$ et N un entier vérifiant la condition (1.13) et t.q. $X_N \in \mathcal{C}$. On définit

$$X'_n = X_n - \sum_{i=n+1}^{\infty} \gamma_i \delta M_i \mathbf{1}_{\{\|X_{i-1} - x^\star\|^2 \leq M+2\}}.$$

En utilisant l'équation (1.11), on peut facilement montrer que X'_n satisfait la relation de récurrence suivante

$$X'_{n+1} = X'_n - \gamma_{n+1} \delta M_{n+1} \mathbf{1}_{\{\|X_n - x^\star\|^2 > M+2\}} - \gamma_{n+1}(u(X_n) - p_{n+1}). \tag{1.14}$$

Nous allons maintenant prouver par récurrence que la suite $(X'_n)_n$ reste dans l'ensemble $\{x : \|x - x^\star\|^2 \leq M + 1\} = \mathcal{C}'$.

L'hypothèse de récurrence est satisfaite pour $n = N$ (c'est suffisant de choisir $\varepsilon < \sqrt{M^2 + 1} - M$). Soit $n > N$, on suppose que l'hypothèse soit satisfaite pour les rangs $N, \ldots, n$, on va montrer qu'elle est encore vraie au rand $n+1$. Alors, $\|X_n - x^\star\|^2 \leq M+2$. On déduit donc de l'équation (1.14) que

$$\begin{aligned} X'_{n+1} &= X'_n - \gamma_{n+1}(u(X_n) - p_{n+1}), \\ \left\|X'_{n+1} - x^\star\right\|^2 &\leq \|X'_n - x^\star\|^2 - 2\gamma_{n+1}(X'_n - x^\star \mid u(X_n)) \\ &\quad + 2\gamma_{n+1}\{\gamma_{n+1}(\|u(X_n)\|^2 + \|p_{n+1}\|^2) + \|p_{n+1}\| \, \|X'_n - x^\star\|\}. \end{aligned} \tag{1.15}$$

En introduisant $c_M = \sup_{\|x - x^\star\|^2 \leq M} \|u(x)\|$, on peut majorer le troisième terme de l'équation (1.15) par $2\varepsilon^2(c_{M+2}^2 + M + 2)$, pourvu que $\varepsilon < 1$.

- Si $\|X'_n - x^\star\|^2 \leq M$, choisir ε tel que $2\varepsilon^2(c_{M+2}^2 + M + 2) + \varepsilon M c_{M+2} \leq 1$ garantit que la somme des deux derniers termes de l'équation (1.15) est plus petite que 1. Ceci assure que $\left\|X'_{n+1} - x^\star\right\|^2 \leq M + 1$.
- Si $M < \|X'_n - x^\star\|^2 \leq M + 1$. Grâce à la continuité de u et à l'hypothèse ($\mathcal{A}$1.4-*i)*, $(X_n - x^\star \mid u(X_n)) > \eta > 0$. On écrit

  $$(X'_n - x^\star \mid u(X_n)) = (X_n - x^\star \mid u(X_n)) + (X'_n - X_n \mid u(X_n)).$$

  Le second terme est majoré en valeur absolue par εc_{M+2}. Ainsi,

  $$(X'_n - x^\star \mid u(X_n)) \leq \eta - \varepsilon c_{M+2}.$$

  En choisissant ε de telle sorte que $\eta - \varepsilon c_{M+2} - \varepsilon(c_{M+2} + M + 2) > 0$, on assure que $\left\|X'_{n+1} - x^\star\right\|^2 < \|X'_n - x^\star\|^2$ et par conséquent $\left\|X'_{n+1} - x^\star\right\|^2 \leq M + 1$.

Nous venons de démontrer que pour tout $n > N$, $\|X'_n - x^\star\|^2 \leq M + 1$. Comme ε peut être choisi plus petit que $\sqrt{M^2+1} - M$, nous avons également le résultat suivant

$$\|X_n - x^\star\|^2 \leq M + 2, \text{ pour tout } n > N.$$

Ceci achève de démontrer que la suite $(X_n)_n$ reste dans un compact et que par conséquent $\limsup_n \sigma_n$ est finie p.s.. ■

1.2.2 Vitesse de convergence

En dépit de l'abondante littérature concernant l'étude asymptotique des algorithmes stochastiques, la vitesse de convergence de l'algorithme aléatoirement tronquée proposée par Chen n'a été que fort peu étudiée en dehors de Pelletier [49] qui démontre un théorème central limite marginal pour le dit algorithme. La démonstration qui en est faite se base sur des arguments de convergence fonctionnelle dans l'espace de Skorokhod.

Nous rappelons ici le résultat obtenu par Pelletier [49, Théorème 1] et nous en donnons une démonstration plus simple (voir paragraphe 1.3) utilisant des arguments de convergence des tableaux triangulaires de martingale. La preuve que nous proposons se base essentiellement sur le théorème central limite pour les martingales (voir page 140 pour un énoncé).

Pourquoi est-ce différent de l'algorithme de Robbins Monro. Il peut être tentant à la vue de la proposition 1.2.1 de vouloir utiliser un argument de type translation du temps et de dire que l'algorithme tronqué de Chen se comporte en fait comme l'algorithme de Robbins Monro à partir d'un certain rang. Ainsi, tous les résultats valables pour l'algorithme de Robbins Monro le seraient également pour l'algorithme tronqué. Malheureusement, un tel argument est rapidement mis en défaut puisque le nombre de troncatures n'est pas borné mais seulement fini presque sûrement et n'est même pas un temps d'arrêt. Cette remarque est tout sauf anodine et c'est bien à cause de cela qu'on ne peut pas déduire de résultats relatifs à la convergence en loi de l'algorithme de Chen à partir de résultats sur l'algorithme de Robbins Monro. En effet, on peut facilement construire des exemples de variables aléatoires $(X_n)_n$ et τ telles que τ soit fini presque sûrement et telles que $(X_n)_n$ converge en loi mais $X_{n+\tau}$ ne converge pas en loi. Il suffit par exemple de considérer τ et τ' 2 v.a. indépendantes de loi de Bernouilli sur $\{0, 1\}$ et de poser $X_n = (-1)^n(\tau - \tau')$. X_n est constante en loi mais mais un calcul de la fonction caractéristique de $X_{n+\tau}$ montre que cette suite translatée ne converge plus.

Les hypothèses utilisées dans cette partie sont très proches de celles décrites pour les algorithmes sous contraintes (voir les hypothèses ($\mathcal{A}1.4$), ($\mathcal{A}1.5$) et ($\mathcal{A}1.8$) page 25 avec $\mathcal{H} = \mathbb{R}$).

Nous avons besoin d'introduire une hypothèse relative à la géométrie de la suite de compacts $(\mathcal{K}_n)_n$. Cette hypothèse est à rapprocher de l'hypothèse ($\mathcal{A}1.6$) pour les algorithmes sous contraintes.

($\mathcal{A}$1.8) Il existe $\eta > 0$ tel que $\forall n \geq 0, \quad d(x^\star, \partial\mathcal{K}_n) > \eta$.

Intéressons-nous maintenant à la vitesse de convergence de la suite $(X_n)_n$ définie par (1.11). Nous rappelons que Δ_n est donnée par

$$\Delta_n = \frac{X_n - x^\star}{\sqrt{\gamma_n}}.$$

Un TCL pour $1/2 < \alpha < 1$

Théorème 1.2.4. *Sous les hypothèses ($\mathcal{A}$1.4), ($\mathcal{A}$1.5) et ($\mathcal{A}$1.8), la suite $(\Delta_n)_n$ converge en loi vers une variable aléatoire de loi normale centrée et de matrice de covariance*

$$V = \int_0^\infty \exp(-At)\Sigma\exp(-At)dt. \tag{1.16}$$

Un TCL pour $\alpha = 1$

Théorème 1.2.5. *Sous les hypothèses ($\mathcal{A}$1.4), ($\mathcal{A}$1.5) et ($\mathcal{A}$1.8) et si de plus $\gamma A - \frac{I}{2}$ est définie positive, la suite $(\Delta_n)_n$ converge en loi vers une variable aléatoire de loi normale centrée et de matrice de covariance*

$$V = \gamma\int_0^\infty \exp\left(\left(\frac{I}{2} - \gamma A\right)t\right)\Sigma\exp\left(\left(\frac{I}{2} - \gamma A\right)t\right)dt. \tag{1.17}$$

Remarque 1.2.6. *Ces théorèmes peuvent être étendus au cas d'hypothèses locales (au voisinage de la solution $\theta^\star$), mais la démonstration s'en trouve de beaucoup compliquée. Nous préférons donc ne la donner que sous des hypothèses globales. Par ailleurs, sous des hypothèses locales, ces deux théorèmes sont en fait des corollaires des théorèmes énoncés page 47.*

1.3 Preuve du TCL pour l'algorithme tronqué de Chen

Des résultats de TCL pour l'algorithme de Robbins Monro existent dans la littérature mais ne sont pas toujours très accessibles. En ce qui concerne l'algorithme de Chen, aucun résultat de vitesse de convergence n'a été démontré en utilisant des arguments de tableaux de martingales, ce qui est pourtant plus lisible que les arguments de convergence fonctionnelle. Dans cette partie, nous allons démontrer les théorèmes 1.2.4 et 1.2.5. Pour rendre la preuve plus accessible, nous l'avons découpée en trois lemmes qui seront démontrés au paragraphe 1.3.2.

1.3.1 Quelques lemmes techniques

Tout d'abord, il nous faut introduire quelques objets supplémentaires. Pour $n > 0$, on définit la fonction $t_n : \mathbb{R}_+ \longrightarrow \mathbb{N}$

$$t_n(u) = \sup \left\{ k \geq 0 \quad ; \quad \sum_{i=n}^{n+k} \gamma_i \leq u \right\}. \tag{1.18}$$

avec la convention $\sup \emptyset = 0$.

On définit $\tilde{\Delta}_n(\cdot)$ comme l'interpolation constante par morceaux de la suite $(\Delta_{n+p})_p$ sur des intervalles de longueur $(\gamma_{n+p})_p$. Plus précisément, on pose

$$\tilde{\Delta}_n(0) = \Delta_n \quad \text{et} \quad \tilde{\Delta}_n(t) = \Delta_{n+t_n(t)+1} \quad \text{pour } t \geq 0. \tag{1.19}$$

Ceci signifie que pour $t \in \left[\sum_{i=n}^{n+p} \gamma_i \ , \ \sum_{i=n}^{n+p+1} \gamma_i\right[$, $t_n(t) = p$ et $\tilde{\Delta}_n(t) = \Delta_{n+p+1}$. $\tilde{X}_n(\cdot)$ est défini de manière similaire.

On introduit également $W_n(.)$

$$W_n(0) = 0 \quad \text{and } W_n(t) = \sum_{i=n+1}^{n+t_n(t)+1} \sqrt{\gamma_i} \delta M_i \quad \text{pour } t > 0. \tag{1.20}$$

Remarque 1.3.1. *Remarquons que les processus $\tilde{\Delta}_n(\cdot)$ et $W_n(\cdot)$ sont des processus càdlàg de saut pur.*

Les théorèmes central limite pour l'algorithme stochastique tronqué (voir Section 1.2.2) sont basés sur les trois lemmes suivants.

Lemme 1.3.2. *Il existe $N_0 > 0$, tel que si on introduit la suite décroissante d'ensembles suivants*

$$A_n = \left\{ \sup_{n \geq m > N_0} \|X_m - x^\star\| < x_0 \right\}, \tag{1.21}$$

on a

$$\sup_{n \geq N_0} \mathbb{E} \left(\|\Delta_n\|^2 \mathbf{1}_{A_n} \right) < \infty. \tag{1.22}$$

De plus, la suite $(\Delta_n)_n$ est tendue.

Remarque 1.3.3. *Remarquons au passage que les ensembles A_n sont mesurables par rapport à la filtration $(\mathcal{F}_n)_n$.*

Lemme 1.3.4. *Il existe 2 suites de processus constants par morceaux $R_n(\cdot)$ et $P_n(\cdot)$ et une matrice Q tels que $\tilde{\Delta}_n(t)$ puisse s'écrire*

$$\begin{aligned} \tilde{\Delta}_n(t) \;\; = \;\; & -\int_0^t e^{Q(u-t)} dW_n(u) + e^{-Qt}\tilde{\Delta}_n(0) - \int_0^t e^{Q(u-t)} y(\tilde{X}_n(u) - x^\star)\tilde{\Delta}_n(u) du \\ & + \int_0^t e^{Q(u-t)} dR_n(u) + \int_0^t e^{Q(u-t)} dP_n(u). \end{aligned} \tag{1.23}$$

Pour tout $n > N_0$ fixé — N_0 étant défini par le lemme 1.3.2 — ,tous les termes de l'égalité précédente excepté l'intégrale stochastique tendent vers zéro en probabilité quand t tend vers l'infini.

Lemme 1.3.5. *Dans (1.23), l'intégrale stochastique $\int_0^t e^{Q(u-t)} dW_n(u)$ converge en loi à $n > N_0$ fixé — N_0 étant défini par le lemme 1.3.2 — vers $\mathcal{N}(0, V)$ quand t tend vers l'infini, où $V = \int_0^\infty e^{-Qs} \Sigma e^{-Qs} ds$.*

Démonstration des théorèmes 1.2.4 et 1.2.5. En utilisant le lemme 1.3.5, il est assez immédiat de voir que pour tout $n > N_0$, $\tilde{\Delta}_n(t)$ converge en loi vers une variable aléatoire de loi normale centrée et de variance V quand t tend vers l'infini. Compte tenu de la définition de $\tilde{\Delta}_n(t)$, il est clair que la convergence de $\tilde{\Delta}_n(t)$ quand t tend vers l'infini implique la convergence de Δ_n vers la même limite quand n tend vers l'infini. En remplaçant la matrice Q par sa valeur en fonction de α, on trouve le résultat annoncé. ■

Remarquons que la preuve pour l'algorithme de Robbins Monro est grandement simplifiée puisqu'il n'est pas nécessaire d'introduire la suite d'ensembles A_n qui permettent ici de traiter les termes dûs aux troncatures successives.

1.3.2 Démonstration des lemmes

Démonstration du lemme 1.3.2

Nous ne faisons la preuve que dans le cas $\alpha = 1$, puisque le schéma de preuve reste identique pour $1/2 < \alpha < 1$, il suffit simplement de modifier quelques développements limités.

Tout d'abord, établissons une relation de récurrence

$$\begin{aligned}
\Delta_{n+1} &= \frac{X_{n+1} - x^\star}{\sqrt{\gamma_{n+1}}}, \\
&= \frac{1}{\sqrt{\gamma_{n+1}}} \left(X_n - x^\star - \gamma_{n+1} u(X_n) - \gamma_{n+1} \delta M_{n+1} + \gamma_{n+1} p_{n+1} \right), \\
&= \sqrt{\frac{\gamma_n}{\gamma_{n+1}}} \, \Delta_n - \sqrt{\gamma_{n+1}} (u(X_n) + \delta M_{n+1} - p_{n+1}).
\end{aligned}$$

Grâce à l'hypothèse ($\mathcal{A}$1.4-*ii)*, l'équation précédente devient

$$\Delta_{n+1} = \left(\sqrt{\frac{\gamma_n}{\gamma_{n+1}}} I - \sqrt{\gamma_{n+1}\gamma_n} A - \sqrt{\gamma_{n+1}\gamma_n} \, y(X_n - x^\star) \right) \Delta_n - \sqrt{\gamma_{n+1}} \delta M_{n+1} + \sqrt{\gamma_{n+1}} p_{n+1}. \tag{1.24}$$

On utilise le développement limité suivant

$$\sqrt{\frac{\gamma_n}{\gamma_{n+1}}} = 1 + \frac{1}{2(n+1)} + \mathcal{O}\left(\frac{1}{n^2}\right) \text{ et } \sqrt{\gamma_n \gamma_{n+1}} = \gamma_n + \mathcal{O}\left(\frac{1}{n^2}\right). \tag{1.25}$$

On pose

$$Q = A - \frac{I}{2\gamma} \tag{1.26}$$

qui est une matrice symétrique définie positive.

Cette remarque nous permet de simplifier l'équation (1.24) en introduisant une nouvelle suite $(\beta_n)_n$ telle que pour tout n plus grand qu'un n_0 fixé, $|\beta_n| \leq C$, où C est une constante strictement positive. L'équation (1.24) se réécrit donc

$$\begin{aligned}\Delta_{n+1} &= \Delta_n - \gamma_n Q\Delta_n - \gamma_n y(X_n - x^\star)\Delta_n - \sqrt{\gamma_{n+1}}\delta M_{n+1} \\ &\quad + \sqrt{\gamma_{n+1}} p_{n+1} + \frac{\beta_n}{(n+1)^2}(B + y(X_n - x^\star))\Delta_n, \end{aligned} \tag{1.27}$$

où B est une matrice déterministe.

Soit $\Delta_{n+\frac{1}{2}} = \dfrac{X_{n+\frac{1}{2}} - x^\star}{\sqrt{\gamma_{n+1}}}$, où $X_{n+\frac{1}{2}}$ est la valeur calculée pour le nouvel itéré avant troncature.

$$\left\|\Delta_{n+\frac{1}{2}}\right\|^2 \leq \left\|\Delta_n - \gamma_n Q\Delta_n - \gamma_n y(X_n - x^\star)\Delta_n - \sqrt{\gamma_{n+1}}\delta M_{n+1} \right. \\ \left. + \frac{\beta_n}{(n+1)^2}(B + y(X_n - x^\star))\Delta_n\right\|^2.$$

En prenant l'espérance conditionnelle par rapport à $\mathcal{F}_n$ (notée $\mathbb{E}_n$) on obtient

$$\mathbb{E}_n\left(\left\|\Delta_{n+\frac{1}{2}}\right\|^2\right) \leq \|\Delta_n\|^2 - 2\gamma_n \Delta_n{}'(Q + y(X_n - x^\star))\Delta_n \\ + \mathcal{O}\left(\frac{1}{n^2}\right)\|\Delta_n\|^2 + \gamma_{n+1}\mathbb{E}_n(\|\delta M_{n+1}\|^2).$$

A ce stade, nous pouvons préciser la définition des ensembles A_n.
Puisque X_n converge presque sûrement vers $x^\star$,

$$\forall \varepsilon > 0,\ \forall \eta > 0, \exists N > 0 \text{ tel que } \forall n \geq N\ \mathbb{P}\left(\sup_{m>n}\|X_m - x^\star\| > \eta\right) < \varepsilon. \tag{1.28}$$

Soit λ la plus petite valeur propre de Q. Comme Q est symétrique définie positive, $\lambda > 0$. $\lim_{\|x\|\to 0} y(x) = 0$, donc pour $x < x_0$, $\|y(x)\| < 3\lambda/4$. Soit $\varepsilon > 0$. Grâce à (1.28), il existe un rang N_0 tel que $\mathbb{P}(\sup_{m>N_0}\|X_m - x^\star\| > x_0) < \varepsilon$.

Dans la définition des ensembles A_n (voir (1.21)), nous choisissons N_0 comme défini ci-dessus et plus grand que n_0.

Sur l'ensemble A_n, $Q+y(X_n-x^\star)$ est une matrice définie positive de plus petite valeur propre plus grande que $\lambda/4$. Par conséquent, $\Delta_n{}'(Q+y(X_n-x^\star))\Delta_n > \lambda/2\,\|\Delta_n\|^2$. Nous pouvons supposer que pour $n > N_0$, $\mathcal{O}(\frac{1}{n^2}) \leq \lambda/4\gamma_n$.

$$\begin{aligned}
\mathbb{E}\left(\left\|\Delta_{n+\frac{1}{2}}\right\|^2 \mathbf{1}_{A_n}\right) - \mathbb{E}\left(\|\Delta_n\|^2 \mathbf{1}_{A_n}\right) &\leq -\gamma_n \frac{\lambda}{2}\mathbb{E}\left(\|\Delta_n\|^2 \mathbf{1}_{A_n}\right) + c\gamma_n, \\
\mathbb{E}\left(\left\|\Delta_{n+\frac{1}{2}}\right\|^2 \mathbf{1}_{A_{n+1}}\right) - \mathbb{E}\left(\|\Delta_n\|^2 \mathbf{1}_{A_n}\right) &\leq -\gamma_n \frac{\lambda}{2}\mathbb{E}\left(\|\Delta_n\|^2 \mathbf{1}_{A_n}\right) + c\gamma_n, \qquad (1.29)
\end{aligned}$$

où $c \in \mathbb{R}^+$.

Nous souhaiterions remplacer $\Delta_{n+\frac{1}{2}}$ par Δ_{n+1}.

$$\begin{aligned}
\|\Delta_{n+1}\|^2 &= \frac{\|X_0 - x^\star\|^2}{\gamma_{n+1}}\mathbf{1}_{\{p_{n+1}\neq 0\}} + \left\|\Delta_{n+\frac{1}{2}}\right\|^2 \mathbf{1}_{\{p_{n+1}=0\}}, \\
\|\Delta_{n+1}\|^2 &\leq \left\|\Delta_{n+\frac{1}{2}}\right\|^2 + \frac{\|X_0 - x^\star\|^2}{\gamma_{n+1}}\mathbf{1}_{\{X_n - \gamma_{n+1}U(X_n, Z_{n+1}) \notin \mathcal{K}_{\sigma_n}\}}.
\end{aligned}$$

En prenant l'espérance conditionnelle par rapport à $\mathcal{F}_n$, on trouve

$$\begin{aligned}
\mathbb{E}_n \|\Delta_{n+1}\|^2 &\leq \mathbb{E}_n \left\|\Delta_{n+\frac{1}{2}}\right\|^2 + \frac{\|X_0 - x^\star\|^2}{\gamma_{n+1}}\mathbb{E}_n\left(\mathbf{1}_{\{X_n - \gamma_{n+1}U(X_n, Z_{n+1}) \notin \mathcal{K}_{\sigma_n}\}}\right), \\
\mathbb{E}_n \|\Delta_{n+1}\|^2 \mathbf{1}_{A_n} &\leq \mathbb{E}_n \left\|\Delta_{n+\frac{1}{2}}\right\|^2 \mathbf{1}_{A_n} + \frac{\|X_0 - x^\star\|^2}{\gamma_{n+1}}\mathbf{1}_{A_n}\mathbb{E}_n\left(\mathbf{1}_{\{X_n - \gamma_{n+1}U(X_n, Z_{n+1}) \notin \mathcal{K}_{\sigma_n}\}}\right),
\end{aligned}$$

$$\begin{aligned}
\mathbb{E}\left(\|\Delta_{n+1}\|^2 \mathbf{1}_{A_{n+1}}\right) &\leq \mathbb{E}\left(\left\|\Delta_{n+\frac{1}{2}}\right\|^2 \mathbf{1}_{A_n}\right) + \\
&\quad \frac{\|X_0 - x^\star\|^2}{\gamma_{n+1}}\mathbb{E}\left(\mathbf{1}_{A_n}\mathbb{E}_n\left(\mathbf{1}_{\{X_n - \gamma_{n+1}U(X_n, Z_{n+1}) \notin \mathcal{K}_{\sigma_n}\}}\right)\right). \qquad (1.30)
\end{aligned}$$

L'espérance conditionnelle de droite peut être réécrite

$$\begin{aligned}
\mathbb{E}_n\left(\mathbf{1}_{\{X_n - \gamma_{n+1}U(X_n, Z_{n+1}) \notin \mathcal{K}_{\sigma_n}\}}\mathbf{1}_{A_n}\right) &\leq \mathbb{P}_n\left(\gamma_{n+1}\|U(X_n, Z_{n+1})\| \geq (X_n, \partial\mathcal{K}_{\sigma_n})\right)\mathbf{1}_{A_n}, \\
&\leq \frac{\gamma_{n+1}^2}{d\left(X_n, \partial\mathcal{K}_{\sigma_n}\right)^2}\mathbb{E}_n\left(\|U(X_n, Z_{n+1})\|^2\right)\mathbf{1}_{A_n}. \qquad (1.31)
\end{aligned}$$

De plus, en utilisant l'inégalité triangulaire on trouve

$$d\left(X_n, \partial\mathcal{K}_{\sigma_n}\right) \geq d\left(x^\star, \partial\mathcal{K}_{\sigma_n}\right) - \|X_n - x^\star\|.$$

Grâce à l'hypothèse ($\mathcal{A}1.8$), $d\left(x^\star, \partial\mathcal{K}_{\sigma_n}\right) < \eta$ et sur A_n, $\|X_n - x^\star\| \leq x_0$. Ainsi,

$$d\left(X_n, \partial\mathcal{K}_{\sigma_n}\right) \geq \eta - x_0.$$

Nous pouvons choisir x_0 plus petit que $\eta/2$ par exemple, de telle sorte que $(\eta - x_0)^2 > \frac{\eta^2}{4}$. Ainsi, l'équation (1.31) devient

$$\mathbb{E}\left(\mathbf{1}_{\{X_n - \gamma_{n+1}U(X_n, Z_{n+1}) \notin \mathcal{K}_{\sigma_n}\}}\mathbf{1}_{A_n}\right) \leq \frac{4\gamma_{n+1}^2}{\eta^2}\mathbb{E}\left(\|U(X_n, Z_{n+1})\|^2 \mathbf{1}_{A_n}\right).$$

En utilisant l'hypothèse ($\mathcal{A}$1.5-*i)* et la continuité de u, on trouve

$$\mathbb{E}\left(\|U(X_n, Z_{n+1})\|^2 \mathbf{1}_{A_n}\right) \leq 2\sup_n \mathbb{E}(\|\delta M_n\|^2) + 2 \sup_{\|x-x^\star\|<x_0} \mathbb{E}(u(x)^2).$$

D'où

$$\mathbb{E}\left(\mathbf{1}_{\{X_n-\gamma_{n+1}U(X_n,Z_{n+1})\notin\mathcal{K}_{\sigma_n}\}}\mathbf{1}_{A_n}\right) \leq c\gamma_{n+1}^2.$$

De l'équation (1.30), nous pouvons déduire

$$\mathbb{E}\left(\|\Delta_{n+1}\|^2 \mathbf{1}_{A_{n+1}}\right) \leq \mathbb{E}\left(\left\|\Delta_{n+\frac{1}{2}}\right\|^2 \mathbf{1}_{A_n}\right) + c\gamma_n. \tag{1.32}$$

En combinant les équations (1.32) et (1.29), nous obtenons

$$\mathbb{E}\left(\|\Delta_{n+1}\|^2 \mathbf{1}_{A_{n+1}}\right) \leq \left(1-\gamma_n\frac{\lambda}{4}\right)\mathbb{E}\left(\|\Delta_n\|^2 \mathbf{1}_{A_n}\right) + c\gamma_n.$$

Soit $\mathcal{I} = \left\{i > N_0 : -\frac{\lambda}{4}\mathbb{E}\left(\|\Delta_i\|^2 \mathbf{1}_{A_i}\right) + c > 0\right\}$, alors

$$\sup_{i\in\mathcal{I}} \mathbb{E}\left(\|\Delta_i\|^2 \mathbf{1}_{A_i}\right) < \frac{4c}{\lambda} < \infty.$$

Pour $i \notin \mathcal{I}$,

$$\mathbb{E}\left(\|\Delta_{i+1}\|^2 \mathbf{1}_{A_{i+1}}\right) - \mathbb{E}\left(\|\Delta_i\|^2 \mathbf{1}_{A_i}\right) \leq 0.$$

Nous allons démontrer par récurrence que $\forall i \geq N_0,\ \mathbb{E}\left(\|\Delta_i\|^2 \mathbf{1}_{A_i}\right) \leq \frac{4c}{\lambda} + \mathbb{E}\left(\|\Delta_{N_0}\|^2 \mathbf{1}_{A_{N_0}}\right)$. Cette relation est évidente pour $i = N_0$. Supposons que l'hypothèse de récurrence soit vraie au rang $i > N_0$. Si $i+1 \in \mathcal{I}$, alors $\mathbb{E}\left(\|\Delta_{i+1}\|^2 \mathbf{1}_{A_{i+1}}\right) \leq \frac{4c}{\lambda}$. Sinon, si $i+1 \notin \mathcal{I}$, $\mathbb{E}\left(\|\Delta_{i+1}\|^2 \mathbf{1}_{A_{i+1}}\right) \leq \mathbb{E}\left(\|\Delta_i\|^2 \mathbf{1}_{A_i}\right)$. L'hypothèse de récurrence permet alors de conclure et le résultat annoncé est démontré.

Ainsi,

$$\sup_n \mathbb{E}\left(\|\Delta_n\|^2 \mathbf{1}_{A_n}\right) < \infty.$$

Finalement, cette relation combinée avec (1.28) prouve la tension de la suite $(\Delta_n)_n$. En effet, soit $M > 0$.

$$\begin{aligned}
\mathbb{P}(\|\Delta_n\| > M) &\leq \mathbb{P}(\|\Delta_n\|(\mathbf{1}_{A_n} + \mathbf{1}_{A_n^c}) > M), \\
&\leq \mathbb{P}(\|\Delta_n\| \mathbf{1}_{A_n} > M/2) + \mathbb{P}(\|\Delta_n\| \mathbf{1}_{A_n^c} > M/2), \\
&\leq 4/M^2\, \mathbb{E}\left(\|\Delta_n\|^2 \mathbf{1}_{A_n}\right) + \mathbb{P}(A_n^c).
\end{aligned} \tag{1.33}$$

Il existe une valeur de M dépendant de ε telle que tous les termes de droite dans (1.33) soient majorés par ε. Ceci prouve la tension de $(\Delta_n)_n$ et achève la démonstration du lemme.

Remarque 1.3.6 (cas $\alpha < 1$). *Cette démonstration reste valable dans le cas $\alpha < 1$, il suffit de modifier les développements limités de l'équation (1.25) comme suit*

$$\sqrt{\frac{\gamma_n}{\gamma_{n+1}}} = 1 + \mathcal{O}\left(\frac{1}{n}\right) \quad \text{et } \sqrt{\gamma_n \gamma_{n+1}} = \gamma_n + \mathcal{O}\left(\frac{1}{n^{1+\alpha}}\right).$$

L'équation (1.27) devient alors

$$\begin{aligned} \Delta_{n+1} &= \Delta_n - \gamma_n Q \Delta_n - \gamma_n y(X_n - x^\star)\Delta_n - \sqrt{\gamma_{n+1}}\delta M_{n+1} \\ &\quad + \sqrt{\gamma_{n+1}} p_{n+1} + \frac{\beta_n}{(n+1)}(B + y(X_n - x^\star))\Delta_n, \end{aligned}$$

où $Q = A$ cette fois, mais reste définie positive.

Démonstration du lemme 1.3.4

Si nous retournons à l'équation (1.27) et que nous la sommons de n — choisi plus grand que N_0 — à $n+p$, nous obtenons

$$\begin{aligned} \Delta_{n+p} &= \Delta_n - \sum_{k=0}^{p-1} \gamma_{n+k}(\; Q \; + y(X_{n+k} - x^\star))\Delta_{n+k} - \sqrt{\gamma_{n+k+1}}\delta M_{n+k+1} \\ &\quad + \sum_{k=0}^{p-1} \sqrt{\gamma_{n+k+1}} p_{n+k+1} + \frac{\beta_{n+k}}{n+k+1}\gamma_{n+k}(B + y(X_{n+k} - x^\star))\Delta_{n+k}. \end{aligned} \tag{1.34}$$

Maintenant, choisissons $u > 0$ tel que $t_n(u) = p - 1$. Puisque $\tilde{X}_n(\cdot)$ est constant par morceaux sur la subdivision définie par la suite $(\gamma_{n+p})_{p\geq 0}$, les sommes discrètes peuvent être interprétées comme des intégrales.

$$\tilde{\Delta}_n(u) = \tilde{\Delta}_n(0) - \int_0^u \left(Q + y(\tilde{X}_n(s) - x^\star)\right) \tilde{\Delta}_n(s) ds - W_n(u) + R_n(u) + P_n(u) \tag{1.35}$$

où

$$\begin{aligned} P_n(u) &= \sum_{k=0}^{t_n(u)} \sqrt{\gamma_{n+k+1}} p_{n+k+1}, \\ R_n(u) &= \sum_{k=0}^{t_n(u)} \frac{\beta_{n+k}}{n+k+1}\gamma_{n+k} \, (B + y(X_{n+k} - x^\star))\Delta_{n+k}. \end{aligned} \tag{1.36}$$

Remarquons que

$$\|R_n(u)\| \;\leq\; \frac{C}{n}\int_0^u \left(1 + \left\|y(\tilde{X}_n(s) - x^\star)\right\|\right) \left\|\tilde{\Delta}_n(s)\right\| ds.$$

Considérons plutôt la forme différentielle de l'équation (1.35)

$$d\tilde{\Delta}_n(u) = -\left(Q + y(\tilde{X}_n(u) - x^\star)\right)\tilde{\Delta}_n(u)du - dW_n(u) + dR_n(u) + dP_n(u). \qquad (1.37)$$

Nous pouvons intégrer l'équation (1.37) pour obtenir l'expression annoncée dans le lemme 1.3.4.

Nous devons encore démontrer que tous les termes autres que l'intégrale stochastique tendent vers zéro en probabilité quand t tend vers l'infini.

Pour traiter le premier terme dans (1.23), il suffit de se rappeler que l'ensemble $\{\tilde{\Delta}_n(0); n \geq 0\}$ est tendu. Puisque Q est définie positive, $e^{-Qt}\tilde{\Delta}_n(0)$ tend vers zéro en probabilité quand t tend l'infini.

En ce qui concerne le second terme, on sait que l'ensemble $\{\tilde{\Delta}_n(u); u < \infty\}$ est tendu. Puisque $\tilde{X}_n(u)$ converge presque sûrement vers $x^\star$ quand u tend vers l'infini, $y(\tilde{X}_n(u) - x^\star)$ tend vers zéro presque sûrement. Ces deux conditions impliquent que $y(\tilde{X}_n(u) - x^\star)\tilde{\Delta}_n(u)$ tend vers zéro en probabilité quand u tend vers l'infini (voir la proposition A.0.5). Ainsi, la proposition A.0.7 assure la convergence de la suite de variables aléatoires $(\int_0^t e^{Q(u-t)}y(\tilde{X}_n(u) - x^\star)\tilde{\Delta}_n(u)du)_t$ vers zéro en probabilité.

Le quatrième terme peut être traité exactement comme précédemment. Puisque $R_n(\cdot)$ est un processus de saut pur[1] avec un nombre fini de sauts sur l'intervalle $[0, t]$, l'intégrale stochastique peut être réécrite comme une somme discrète

$$\int_0^t e^{Q(u-t)}dR_n(u) = \sum_{u \leq t} e^{Q(u-t)}\Delta R_n(u). \qquad (1.38)$$

En utilisant (1.36), l'équation précédente se développe ainsi

$$\int_0^t e^{Q(u-t)}dR_n(u) \quad = \quad \sum_{u\leq t} e^{Q(u-t)}(1 + y(\tilde{X}_n(u) - x^\star))\tilde{\Delta}_n(u)\frac{\beta_{n+t_n(u)}}{(n + t_n(u))}\gamma_{n+t_n(u)}.$$

La somme discrète se comporte comme l'intégrale suivante

$$\int_0^t e^{Q(u-t)}(1 + y(\tilde{X}_n(u) - x^\star))\tilde{\Delta}_n(u)c_n(u)du, \qquad (1.39)$$

où la fonction $c_n(\cdot)$ tend vers 0 quand n tend vers l'infini puisque la suite $(\beta_n)_n$ est bornée. Nous pouvons reproduire le même raisonnement pour le second terme de (1.23) pour prouver que la suite $\left(\int_0^t e^{Q(u-t)}(1 + y(\tilde{X}_n(u) - x^\star))\tilde{\Delta}_n(u)c_n(u)du\right)_t$ est tendue. De plus, $((1+y(\tilde{X}_n(u)-x^\star))\tilde{\Delta}_n(u))_u$ est aussi tendue et $(c_n(u))_u$ converge presque sûrement vers zéro. Par conséquent en utilisant la proposition A.0.5, $((1+y(\tilde{X}_n(u)-x^\star))\tilde{\Delta}_n(u)c_n(u))_u$

[1]Pour plus de détails sur les processus de saut pur et les intégrales stochastiques par rapport à une semi-martingale, nous renvoyons le lecteur à Rogers et Williams [56].

tend vers zéro en probabilité. La proposition A.0.7 permet de conclure que le quatrième terme dans (1.35) tend également vers zéro en probabilité quand t tend vers l'infini.

Le terme dû aux troncatures peut se réécrire

$$\int_0^t e^{Q(u-t)} dP_n(u) = \sum_{i=0}^{t_n(t)} e^{Q(t_n^{-1}(i)-t)} \sqrt{\gamma_{i+n}}\, p_{i+n}. \tag{1.40}$$

La somme discrète se comporte comme $\int_0^t e^{Q(u-t)} \sqrt{\gamma_{n+t_n(u)}}\, p_{n+t_n(u)} du$. Puisqu'il y a un nombre fini presque sûrement de troncatures, p_n est presque sûrement nul pour n assez grand. Par conséquent $\sqrt{\gamma_{n+t_n(u)}}\, p_{n+t_n(u)}$ converge vers zéro presque sûrement quand u tend vers l'infini. La proposition A.0.6 permet alors de prouver que la somme dans (1.40) converge vers zéro presque sûrement quand t tend vers l'infini.

Démonstration du lemme 1.3.5

En utilisant le théorème A.0.8, nous allons démontrer le lemme 1.3.5.

Démonstration. Dans un souci de clarté, nous ferons la preuve en supposant que Q est une constante réelle strictement positive et que la suite de processus $(W_n(\cdot))_n$ est à valeurs réelles.

Nous définissons N_l^p pour tout $0 \le l \le p$ et $p > 0$ par

$$N_l^p = \sum_{i=1}^{l} e^{Q(t_n^{-1}(i)-t_n^{-1}(p))} \sqrt{\gamma_{i+n}} \delta M_{i+n}. \tag{1.41}$$

A p fixé, $(N_l^p)_{0 \le l \le p}$ est évidemment une martingale pour la filtration $(\mathcal{F}_{n+l})_l$ et satisfait la relation $N_p^p = \int_0^{t_n^{-1}(p)} e^{Q(u-t)} dW_n(u)$. Ainsi, nous devons uniquement prouver que N_p^p converge vers une variable aléatoire de loi normale quand p tend l'infini.

Calculons le crochet de la martingale N.

$$\langle N \rangle_p^p \;=\; \sum_{i=1}^{p} e^{2Q(t_n^{-1}(i)-t_n^{-1}(p))} \gamma_{i+n} \mathbb{E}\left(\delta M_{i+n}^2 | \mathcal{F}_{n+i-1}\right). \tag{1.42}$$

Grâce à l'hypothèse ($\mathcal{A}$1.5-*ii)*, l'espérance conditionnelle dans (1.42) converge en probabilité vers Σ quand n tend vers l'infini. De plus, $\langle N \rangle_p^p$ se comporte comme

$$\int_0^{t_n^{-1}(p)} e^{2Q(u-t_n^{-1}(p))} \mathbb{E}\left(\delta M_{t_n(u)+n}^2 | \mathcal{F}_{n+t_n(u)-1}\right) du. \tag{1.43}$$

Nous venons de voir que l'espérance conditionnelle ci-dessus converge en probabilité vers Σ. Par la proposition A.0.7, $\left(\int_0^t e^{2Q(u-t)} \mathbb{E}\left(\delta M_{t_n(u)+n}^2 | \mathcal{F}_{n+t_n(u)-1}\right) du\right)_t$ converge en probabilité vers $\frac{\Sigma}{2Q}$ quand t tend vers l'infini. De plus $t_n^{-1}(p)$ tend vers l'infini quand p

tend vers l'infini, d'où $\langle N\rangle_p^p \xrightarrow[p\to\infty]{\mathbb{P}} \frac{\Sigma}{2Q}$.
Soit ρ le nombre réel défini dans le théorème 1.2.4.

$$\sum_{l=1}^{p} \mathbb{E}\left(\left\|N_l^{(p)} - N_{l-1}^{(p)}\right\|^{2+\rho}\right) = \sum_{i=1}^{p} e^{(2+\rho)Q(t_n^{-1}(i)-t_n^{-1}(p))}\gamma_{i+n}^{1+\frac{\rho}{2}}\mathbb{E}\left(\|\delta M_{i+n}\|^{2+\rho}\right).$$

Cette somme se comporte comme sa version continue

$$\int_0^{t_n^{-1}(p)} e^{(2+\rho)Q(u-t_n^{-1}(p))}\gamma_{t_n(u)+n}^{\frac{\rho}{2}}\mathbb{E}\left(\left\|\delta M_{t_n(u)+n}\right\|^{2+\rho}\right) du. \tag{1.44}$$

$\gamma_{t_n(u)+n}^{\frac{\rho}{2}}$ converge vers 0 quand u tend vers l'infini et la suite des espérances conditionnelles est bornée grâce l'hypothèse ($\mathcal{A}$1.5-*i*). Ainsi $\gamma_{t_n(u)+n}^{\frac{\rho}{2}}\mathbb{E}\left(\left\|\delta M_{t_n(u)+n}\right\|^{2+\rho}\right)$ tend vers zéro quand u tend vers l'infini. La proposition A.0.6 prouve que l'intégrale dans l'équation (1.44) tend vers 0 quand p tend vers l'infini. Donc, $\sum_{l=1}^{p} \mathbb{E}\left(\left\|N_l^{(p)} - N_{l-1}^{(p)}\right\|^{2+\rho} \middle| \mathcal{F}_{l-1}^{(p)}\right)$ tend vers zéro dans $\mathbb{L}^1$, et par conséquent en probabilité. Les hypothèses du théorème A.0.8 sont donc satisfaites. Ceci termine de prouver que l'intégrale stochastique $\int_0^t e^{Q(u-t)}dW_n(u)$ converge en loi vers une variable aléatoire de loi normale centrée et de variance $\frac{\Sigma}{2Q}$. ■

Chapter 2

A CLT for averaging and truncated stochastic algorithms

Abstract

This article is devoted to the study of the convergence rate of the averaging version of the randomly truncated stochastic algorithm introduced by Chen and Zhu [21]. This result is proved by establishing the convergence in the Skorokhod space of a well defined interpolation of the renormalised iterates to a stationary Ornstein Uhlenbeck process.

2.1 Introduction

Stochastic algorithms in the spirit of the Robbins-Monro algorithm are commonly used for solving challenging optimisation problems. These procedures are especially efficient when the function to be minimised is defined as an expectation. Note, for instance, that variance minimisation for Monte Carlo procedures may lead to this kind of problems (see Arouna [7] for an example in a financial context). Unfortunately, the convergence of these algorithms is hung up to assumptions that are barely satisfied in practice, namely the sub-linear growth of the criteria. Some improvements are required for a practical usage.

First, it is necessary to be able to deal with fast growing functions, which has led to introducing truncating techniques. The basic idea is to prevent the algorithm from blowing up during the first steps by forcing the iterates to remain in an increasing sequence of compact sets. This procedure, known as a randomly truncated stochastic algorithm, was first introduced by Chen and Zhu [21]. This modification of the standard procedure is often needed in applications. In the financial example quoted above, because the payoffs involved are completely non-linear, the standard algorithm quickly fails and the truncating technique is unavoidable.

Secondly, it is often wise to use an averaging procedure to smooth the numerical behaviour of the algorithm and to ease the tuning of the step parameter which is known to monitor the numerical efficiency of the algorithm. Averaging algorithms have already been studied by Polyak and Juditsky [50], Kushner and Yang [43], but not in combination with random truncations. For the averaging standard Robbins Monro algorithm, Gahbiche and Pelletier [33] proposed a way to estimate the asymptotic covariance matrix.

The purpose of this article is to prove a convergence rate result for the algorithm with both random truncations and averaging. The almost sure convergence of this combined procedure is a clear consequence of Chen's results (see Chen et al. [22] for the reference paper on this subject and Delyon [28] for an alternative proof). The use of this combined procedure will be justified hereafter by the study of its convergence rate.

An introductory approach to the convergence rate of the Robbins Monro algorithm can be found in Duflo [29]. Bouton [18] and then Benveniste et al. [13] have established functional Central Limit Theorems for the same algorithm. A more elaborate algorithm has been proposed by Dupuis and Kushner [30] and Buche and Kushner [19] for problems to be solved under constraints. They have considered a projected version of the standard algorithm onto a fixed compact set and have proved a functional CLT for this constrained algorithm. From that functional result, Kushner and Yin [44] could derive a CLT for averaging constrained algorithms. The component-wise convergence rate of the randomly truncated algorithm can be deduced from Pelletier [49]. No result for averaging and randomly truncated algorithm exists. Note that Chen et al. [22] proved that the number of truncations in the algorithm is a.s. finite. But unfortunately, the number of truncations being unbounded, the CLT cannot be derived as a trivial consequence of this fact[1].

The main originality of the work lies in the combination of random truncations with averaging. The combined algorithm we study here has the double advantage to deal with fast growing functions and to smooth out the numerical behaviour of the algorithm whereas these are precisely the two grievances people often show against stochastic algorithms. Note that in the non averaging case, our result improves Pelletier's one as we managed to remove the assumption on the truncating term. This assumption was hard to check in practice whereas our local assumptions are easier to satisfy since roughly speaking it is sufficient to verify that the criteria is uniformly squared integrable on compact sets. Under these local assumptions, we prove a functional CLT for the randomly truncated algorithm and a component-wise CLT for the averaging and randomly truncated algorithm.

Here is the outline of the paper. In Section 2.2, we present the framework and our main result: a Central Limit Theorem for averaging and randomly truncated algorithms. This result relies on a functional CLT for the non averaging version of the said algorithm. This functional result is also stated in Section 2.2. Finally, Section 2.3 is devoted to the proof of the functional CLT for non averaging and randomly truncated algorithms.

[1] see the explanation in Section 1.2.2

2.2 CLT for averaging and randomly truncated procedures

2.2.1 Notations and assumptions

Let $u\colon X \in \mathbb{R}^d \longmapsto u(X) \in \mathbb{R}^d$ be a continuous function defined as an expectation on a probability space $(\Omega, \mathcal{A}, \mathbb{P})$.

$$u(X) = \mathbb{E}(U(X, Z)),$$

where Z is a random variable in $\mathbb{R}^m$ and U a measurable function defined on $\mathbb{R}^d \times \mathbb{R}^m$. We assume that $x^\star$ is the unique solution of $u(X) = 0$.

Chen and Zhu [21] introduce a new procedure to approximate $x^\star$. This procedure enables to monitor the excursions of the approximating sequence $(X_n)_n$ outside an increasing sequence of compact sets $(\mathcal{K}_j)_j$ of $\mathbb{R}^d$

$$\bigcup_{j=0}^{\infty} \mathcal{K}_j \;=\; \mathbb{R}^d \quad \text{and} \quad K_j \subsetneq \text{int}(K_{j+1})$$

where $\text{int}(A)$ denotes the interior of the set A.

For $X_0 \in \mathcal{K}_0$ and $\sigma_0 = 0$, we define the sequences of random variables $(X_n)_n$ and $(\sigma_n)_n$

$$\begin{cases} & X_{n+\frac{1}{2}} = X_n - \gamma_{n+1} U(X_n, Z_{n+1}), \\ \text{if } X_{n+\frac{1}{2}} \in \mathcal{K}_{\sigma_n} & X_{n+1} = X_{n+\frac{1}{2}} \quad \text{and} \quad \sigma_{n+1} = \sigma_n, \\ \text{if } X_{n+\frac{1}{2}} \notin \mathcal{K}_{\sigma_n} & X_{n+1} = X_0 \quad \text{and} \quad \sigma_{n+1} = \sigma_n + 1. \end{cases} \tag{2.1}$$

where $(Z_n)_n$ an i.i.d. sequence of random variables following the law of Z and $\gamma_n = \frac{\gamma}{(n+1)^\alpha}$, with $1/2 < \alpha < 1$. $X_{n+\frac{1}{2}}$ is the new sample we draw, either we accept it and set $X_{n+1} = X_{n+\frac{1}{2}}$ or we reject it and reset the algorithm to X_0.

Remark 2.2.1. *When $X_{n+\frac{1}{2}} \notin \mathcal{K}_{\sigma_n}$, one can set X_{n+1} to any measurable function of $(X_0, \ldots, X_n)$ with values in a given compact set. This existence of such a compact set is definitely essential to proof the a.s. convergence of $(X_n)_n$.*

We introduce $\mathcal{F}_n = \sigma(Z_k; k \leq n)$ the σ-field generated by the random vectors Z_k, for $k \leq n$. Note that X_n is $\mathcal{F}_n-$measurable, hence we can write $u(X_n) = \mathbb{E}[U(X_n, Z_{n+1}) | \mathcal{F}_n]$.

Based on this algorithm, we can introduce an averaging algorithm. For any $t > 0$, we define a moving window average of the iterates

$$\hat{X}_n(t) = \frac{\gamma_n}{t} \sum_{i=n}^{n+\lfloor t/\gamma_n \rfloor} X_i. \tag{2.2}$$

We are interested in the convergence of the renormalised iterates of Equation (2.2) centred about its limit

$$\hat{\Delta}_n(t) = \frac{\hat{X}_n(t) - x^\star}{\sqrt{\gamma_n}}.$$

In order to give a more unified presentation, we prefer to rewrite Algorithm (2.1) as

$$X_{n+1} = X_n - \gamma_{n+1} u(X_n) - \gamma_{n+1}\delta M_{n+1} + \gamma_{n+1} p_{n+1} \tag{2.3}$$

where

$$\begin{aligned} \delta M_{n+1} &= U(X_n, Z_{n+1}) - u(X_n), \\ \text{and} \quad p_{n+1} &= \begin{cases} u(X_n) + \delta M_{n+1} + \frac{1}{\gamma_{n+1}}(X_0 - X_n) & \text{if } X_{n+\frac{1}{2}} \notin \mathcal{K}_{\sigma_n}, \\ 0 & \text{otherwise.} \end{cases} \end{aligned}$$

In the following, the prime notation stands for the transpose operator. $(\cdot|\cdot)$ denotes the standard Euclidean scalar product on $\mathbb{R}^d$. To prove a convergence rate result for Algorithm (2.3), we need to introduce three kinds of hypotheses.

($\mathcal{A}$2.1) *i.* $\forall X \in \mathbb{R}^d,\ X \neq x^\star,\ (X - x^\star | u(X)) > 0.$

ii. There exist a function $y : \mathbb{R}^d \to \mathbb{R}^{d\times d}$ satisfying $\lim_{\|x\|\to 0} \|y(x)\| = 0$ and a symmetric definite positive matrix A such that

$$u(X) = A(X - x^\star) + y(X - x^\star)(X - x^\star).$$

($\mathcal{A}$2.2) *i.* There exist two real numbers $\rho > 0$ and $\eta > 0$ such that

$$\kappa = \sup_n \mathbb{E}\left(\|\delta M_n\|^{2+\rho} \mathbf{1}_{\{\|X_{n-1} - x^\star\| \le \eta\}}\right) < \infty.$$

we set $\kappa_0 = \sup_n \mathbb{E}\left(\|\delta M_n\|^2 \mathbf{1}_{\{\|X_{n-1} - x^\star\| \le \eta\}}\right)$.

ii. There exists a symmetric definite positive matrix Σ such that

$$\mathbb{E}\left(\delta M_n \delta M_n' | \mathcal{F}_{n-1}\right) \mathbf{1}_{\{\|X_{n-1} - x^\star\| \le \eta\}} \xrightarrow[n\to\infty]{\mathbb{P}} \Sigma.$$

($\mathcal{A}$2.3) There exists $\mu > 0$ such that $\forall n \ge 0 \quad d(x^\star, \partial\mathcal{K}_n) > \mu$.

Remark 2.2.2. *Comments on the assumptions.*

1. *Hypothesis ($\mathcal{A}$2.1-i) is satisfied as soon as u can be interpreted as the gradient of a strictly convex function. The second point of Hypothesis ($\mathcal{A}$2.1) is equivalent to saying that u is differentiable at the point $x^\star$ with derivative A.*

2. *Hypothesis ($\mathcal{A}$2.2) corresponds to some local uniform integrability property for the family of r.v. $\left(\|U(X, Z_{n+1})\|^2\right)_n$ when X is a compact neighbourhood of $x^\star$.*

3. *Hypothesis ($\mathcal{A}$2.3) is only required for technical reasons but one does not need to be concerned with it in practical situations.*

2.2.2 Main result

Theorem 2.2.3, which is our main result, states a convergence rate for averaging and randomly truncated algorithm. This result highly relies on Theorem 2.2.4 which gives a functional result for the asymptotic behaviour of the non-averaging version of the randomly truncated algorithm. The proof of Theorem 2.2.4 is postponed to Section 2.3.

Theorem 2.2.3. *Under Hypotheses ($\mathcal{A}2.1$) to ($\mathcal{A}2.3$), the sequence $\hat{\Delta}_n(t)$ converges in distribution to a normally distributed random variable with mean* 0 *and variance*

$$\hat{V}(t) = \frac{1}{t}A^{-1}\Sigma A^{-1} + \frac{A^{-2}(e^{-At} - I)V + VA^{-2}(e^{-At} - I)}{t^2}$$

where $V = \int_0^\infty e^{-Au}\Sigma e^{-Au}du$.

We now need to introduce a few more notations needed in the proof of Theorem 2.2.3. We define the sequence of the renormalised iterates of Algorithm (2.3)

$$\Delta_n = \frac{X_n - x^\star}{\sqrt{\gamma_n}}, \quad \text{for all } n \geq 0.$$

We now introduce a sequence of interpolating times $\{t_n(u); u \geq 0, n \geq 0\}$

$$t_n(u) = \sup\left\{k \geq 0 \quad ; \quad \sum_{i=n}^{n+k} \gamma_i \leq u\right\}. \tag{2.4}$$

with the convention $\sup \emptyset = 0$.

We define $\Delta_n(\cdot)$ as the piecewise constant interpolation of $(\Delta_{n+p})_p$ on intervals of length $(\gamma_{n+p})_p$. More precisely,

$$\Delta_n(0) = \Delta_n \quad \text{and} \quad \Delta_n(t) = \Delta_{n+t_n(t)} \quad \text{for } t > 0. \tag{2.5}$$

This means that for $t \in \left[\sum_{i=n}^{n+p} \gamma_i \ , \ \sum_{i=n}^{n+p+1} \gamma_i\right[$, $t_n(t) = p$ and $\Delta_n(t) = \Delta_{n+p+1}$. $X_n(\cdot)$ is defined similarly.

We also introduce $W_n(.)$

$$W_n(0) = 0 \quad \text{and} \quad W_n(t) = \sum_{i=n+1}^{n+t_n(t)+1} \sqrt{\gamma_i}\delta M_i \quad \text{for } t > 0. \tag{2.6}$$

The following theorem will play a key role in the proof of Theorem 2.2.3.

Theorem 2.2.4. *If we assume Hypotheses ($\mathcal{A}2.1$)-($\mathcal{A}2.3$), the sequence of processes $(\Delta_n(\cdot))_n$ converges in law to a diffusion $\Delta(\cdot)$ satisfying*

$$\Delta(t) = \Delta(0)e^{-At} + \int_0^t e^{A(u-t)}dW(u),$$

where $\Delta(0)$ is a random normal variable with mean 0 and variance

$$V = \int_0^\infty e^{-Au}\Sigma e^{-Au}du$$

and W is a Wiener process w.r.t. the smallest $\sigma-$algebra that measures $(\Delta(\cdot), W(\cdot))$ with covariance matrix Σ.

Remark 2.2.5. *We say that a sequence of càdlàg processes X_n converges in law to X or weakly converges in $\mathbb{D}([0,T])$ ($X_n \Longrightarrow X$) if $\mathcal{L}(X_n) \to \mathcal{L}(X)$ weakly in the set of all probability measures defined on $\mathbb{D}([0,T])$, where $\mathbb{D}([0,T])$ is the space of processes defined on $[0,T]$ with almost sure right-continuous paths, left-hand limits and values in $\mathbb{R}^d$. Since the limits involved do not depend on T, the space $\mathbb{D}([0,T])$ will simply be denoted $\mathbb{D}$. One can refer to Billingsley [14] or Jacod and Shiryaev [40] for more details on the weak convergence of càdlàg processes.*

Proof of Theorem 2.2.3. We introduce

$$\widetilde{\Delta}_n(t) = \frac{1}{t}\sum_{i=n}^{n+t_n(t)-1}(X_i - x^\star)\sqrt{\gamma_n}$$

Since $\gamma_n = \frac{\gamma}{n^\alpha}$ with $\frac{1}{2} < \alpha < 1$,

$$t \geq \frac{t_n(t)}{(n+t_n(t))^\alpha}.$$

It is easy to deduce from this inequality that $\frac{t_n(t)}{n} \underset{n}{\to} 0$.

$$\begin{aligned}
\widetilde{\Delta}_n(t) &= \frac{1}{t}\sum_{i=n}^{n+t_n(t)-1}\Delta_i\gamma_i\frac{\sqrt{\gamma_n}-\sqrt{\gamma_i}}{\sqrt{\gamma_i}} + \frac{1}{t}\sum_{i=n}^{n+t_n(t)-1}\Delta_i\gamma_i \\
&= \frac{1}{t}\int_0^t \Delta_n(u)\frac{\sqrt{\gamma_n}-\sqrt{\gamma_{n+t_n(u)}}}{\sqrt{\gamma_{n+t_n(u)}}}du + \frac{1}{t}\int_0^t \Delta_n(u)du \qquad (2.7)
\end{aligned}$$

Thanks to Theorem 2.2.4, $\Delta_n(\cdot) \Rightarrow \Delta(\cdot)$ in $\mathbb{D}([0,T])$ for all $T > 0$. Hence, the second term in (2.7) converges in distribution to $\frac{1}{t}\int_0^t \Delta(u)du$. Moreover, by the fact that

$$\frac{\sqrt{\gamma_n}-\sqrt{\gamma_{n+t_n(t)}}}{\sqrt{\gamma_{n+t_n(t)}}} \underset{n}{\to} 0,$$

the first term in (2.7) tends to zero in probability.

$\Delta(u)du$ is a normally distributed random variable with mean 0 and variance $\hat{V}$.

$$\begin{aligned}
\hat{V} &= \frac{1}{t^2}\operatorname{Cov}\left(\int_0^t \Delta(u)du, \int_0^t \Delta(s)ds\right), \\
&= \frac{1}{t^2}\int_0^t\int_0^t \operatorname{Cov}(\Delta(u),\Delta(s))du\,ds,
\end{aligned}$$

Before diving into the computation, we make a few remarks. Thanks to the definition of the process $\Delta(\cdot)$, it is easy to show that

$$\begin{aligned} \mathrm{Cov}(\Delta(s+\tau),\Delta(s)) &= e^{-A\tau}\,\mathrm{Cov}(\Delta(s),\Delta(s)) && (2.8a)\\ \mathrm{Cov}(\Delta(s),\Delta(s+\tau)) &= \mathrm{Cov}(\Delta(s),\Delta(s))\,\mathrm{e}^{-A\tau}\,. && (2.8b)\end{aligned}$$

Since $\Delta(\cdot)$ is a stationary process, $\mathrm{Cov}(\Delta(s),\Delta(s)) = \mathrm{Cov}(\Delta(0),\Delta(0)) = V$ for any $s \geq 0$. Henceforth, Equations (2.8a) and (2.8b) can be rewritten

$$\mathrm{Cov}(\Delta(s+\tau),\Delta(s)) = e^{-A\tau}V \quad \text{and} \quad \mathrm{Cov}(\Delta(s),\Delta(s+\tau)) = V\,\mathrm{e}^{-A\tau}\,.$$

Let us go back to the computation of $\hat{V}$. Since Cov is a bilinear operator,

$$\begin{aligned} \hat{V} &= \frac{1}{t^2}\left(\int_0^t\int_0^t \mathrm{Cov}(\Delta(u),\Delta(s))\mathbf{1}_{\{u\leq v\}}du\,dv + \int_0^t\int_0^t \mathrm{Cov}(\Delta(u),\Delta(s))\mathbf{1}_{\{v\leq u\}}du\,dv\right),\\ &= \frac{1}{t^2}\left(\int_0^t\int_0^v V\,\mathrm{e}^{-A(v-u)}\,du\,dv + \int_0^t\int_0^u \mathrm{e}^{-A(u-v)}\,V dv\,du\right),\\ &= \frac{1}{t^2}\left\{V\left(A^{-1}t + A^{-2}[\mathrm{e}^{-At} - I]\right) + \left(A^{-1}t + A^{-2}[\mathrm{e}^{-At} - I]\right)V\right\},\\ &= \frac{1}{t}\left(VA^{-1} + A^{-1}V\right) + \frac{1}{t^2}\left\{V\left(A^{-2}[\mathrm{e}^{-At} - I]\right) + \left(A^{-2}[\mathrm{e}^{-At} - I]\right)V\right\}. \qquad (2.9)\end{aligned}$$

Moreover considering the definition of $V = \int_0^\infty \mathrm{e}^{-Au}\,\Sigma\,\mathrm{e}^{-Au}\ du$, it is quite easy to show that V solves the following Ricatti equation

$$AV + VA = \Sigma. \qquad (2.10)$$

One can even prove that V is the unique solution of (2.10). From (2.10), one can deduce that $A^{-1}V + VA^{-1} = A^{-1}\Sigma A^{-1}$. Plugging this last result back into (2.9) gives value of the variance announced in Theorem 2.2.3.

We still have to prove that $\hat{\Delta}_n(t)$ and $\tilde{\Delta}_n(t)$ have the same limit for each $t > 0$.

$$\hat{\Delta}_n(t) - \tilde{\Delta}_n(t) = \frac{\sqrt{\gamma_n}}{t}\sum_{i\in(t_n(t),\lfloor t/\gamma_n\rfloor)}(X_i - x^\star). \qquad (2.11)$$

Using (2.4),

$$t_n(t)\gamma_n \geq t \geq t_n(t)\gamma_n\left(1 + \frac{t_n(t)}{n}\right)^{-\alpha}. \qquad (2.12)$$

A simple Taylor expansion shows that $t/\gamma_n - t_n(t)$ tends to zero when n goes to infinity. Hence, the number of terms in the sum of (2.11) goes to zero and X_i converges a.s. to $x^\star$. Finally, $\hat{\Delta}_n(t) - \tilde{\Delta}_n(t)$ goes to zero a.s. as n goes to infinity, this yields the desired conclusion. ■

2.3 Proof of Theorem 2.2.4

The proof of Theorem 2.2.4 will be achieved throughout a series of lemmas whose proofs — mainly based on tightness criteria in Skorokhod's space — are postponed to Section 2.3.2.

2.3.1 Technical lemmas

Lemma 2.3.1. *There exists $N_0 > 0$ such that if we introduce the following sequence of increasing sets*

$$\mathcal{A}_n = \left\{ \sup_{n \geq m > N_0} \|X_m - x^\star\| < x_0 \right\}, \tag{2.13}$$

we have

$$\sup_{n \geq N_0} \mathbb{E}\left(\|\Delta_n\|^2 \mathbf{1}_{\{\mathcal{A}_n\}} \right) < \infty. \tag{2.14}$$

Moreover, the sequence $(\Delta_n)_n$ is tight in $\mathbb{R}^d$.

Remark 2.3.2. *Note that $(\mathcal{A}_n)_n$ is a decreasing sequence of measurable sets w.r.t $(\mathcal{F}_n)_n$.*

Lemma 2.3.3. *For the value of N_0 introduced in Lemma 2.3.1, we have for any $T > 0$*

$$\sup_{n \geq N_0} \mathbb{E}\left(\sup_{t \leq T} \|\Delta_n(t)\|^2 \mathbf{1}_{\{\mathcal{A}_{n+t_n(T)}\}} \right) < \infty.$$

Moreover, the sequence $\left(\sup_{t \leq T} \|\Delta_n(t)\|^2\right)_n$ is tight in $\mathbb{R}^+$.

Lemma 2.3.4. *$(W_n(t))_{0 \leq t \leq T}$ converges in law to a process W, which is a Wiener process w.r.t. the filtration it generates with covariance matrix Σ.*

Lemma 2.3.5 (Aldous' criteria)**.** *For any positive μ and ε, there exists $0 < \delta < 1$ such that we have the following inequality*

$$\limsup_n \sup \left\{ \mathbb{P}\left(\|\Delta_n(\tau) - \Delta_n(S)\| \geq \mu\right); \begin{array}{c} S \text{ and } \tau \text{ stopping times in } [0,T], \\ S \leq \tau \leq (S+\delta) \wedge T \end{array} \right\} \leq \varepsilon. \tag{2.15}$$

Lemma 2.3.6. *$(W_n(\cdot), \Delta_n(\cdot))_n$ is tight in $\mathbb{D} \times \mathbb{D}$ and converges in law to (W, Δ) where W is a Wiener process with respect to the smallest $\sigma-$algebra that measures $(W(\cdot), \Delta(\cdot))$ with covariance matrix Σ and Δ is the stationary solution of*

$$d\Delta(t) = -A\Delta(t)dt - dW(t).$$

2.3.2 Proofs of the Lemmas

Before proving the different Lemmas, we need a result stating the almost sure convergence of the sequence $(X_n)_n$. A proof of the following Proposition can be found in Chen and Zhu [21] or Delyon [28].

The almost sure convergence of $(X_n)_n$ to $x^\star$ is established in Proposition 1.2.1 which also states that the sequence $(\sigma_n)_n$ is a.s. finite (i.e. for n large enough $p_n = 0$ a.s.).

Proof of Lemma 2.3.1

First, we establish a recursive relation

$$\begin{aligned}\Delta_{n+1} &= \frac{X_{n+1}-x^\star}{\sqrt{\gamma_{n+1}}},\\ &= \frac{1}{\sqrt{\gamma_{n+1}}}\left(X_n - x^\star - \gamma_{n+1}u(X_n) - \gamma_{n+1}\delta M_{n+1} + \gamma_{n+1}p_{n+1}\right),\\ &= \sqrt{\frac{\gamma_n}{\gamma_{n+1}}}\,\Delta_n - \sqrt{\gamma_{n+1}}(u(X_n)+\delta M_{n+1} - p_{n+1}).\end{aligned}$$

Using Hypothesis ($\mathcal{A}2.1$), the previous equation becomes

$$\Delta_{n+1} = \left(\sqrt{\frac{\gamma_n}{\gamma_{n+1}}}I - \sqrt{\gamma_{n+1}\gamma_n}(A+y(X_n-x^\star))\right)\Delta_n - \sqrt{\gamma_{n+1}}(\delta M_{n+1}+p_{n+1}). \quad (2.16)$$

The following Taylor expansions hold

$$\sqrt{\frac{\gamma_n}{\gamma_{n+1}}} = 1 + \mathcal{O}\left(\frac{1}{n}\right) \text{ and } \sqrt{\gamma_n\gamma_{n+1}} = \gamma_n + \mathcal{O}\left(\frac{1}{n^{1+\alpha}}\right). \quad (2.17)$$

We define

This remark enables us to simplify Equation (2.16) by introducing a new sequence $(\beta_n)_n$ such that for any n larger than some fixed n_0, $|\beta_n| \le C$, where C is a positive real constant. Equation (2.16) can be rewritten as

$$\begin{aligned}\Delta_{n+1} &= \Delta_n - \gamma_n A\Delta_n - \gamma_n y(X_n-x^\star)\Delta_n - \sqrt{\gamma_{n+1}}\delta M_{n+1}\\ &\quad +\sqrt{\gamma_{n+1}}p_{n+1} + \frac{\beta_n}{(n+1)}(B+y(X_n-x^\star))\Delta_n, \end{aligned} \quad (2.18)$$

where B is a deterministic matrix.

Let $\Delta_{n+\frac{1}{2}} = \dfrac{X_{n+\frac{1}{2}}-x^\star}{\sqrt{\gamma_{n+1}}}$.

$$\left\|\Delta_{n+\frac{1}{2}}\right\|^2 \le \left\|\Delta_n - \gamma_n(Q+y(X_n-x^\star))\Delta_n - \sqrt{\gamma_{n+1}}\delta M_{n+1} + \frac{\beta_n}{n+1}(B+y(X_n-x^\star))\Delta_n\right\|^2.$$

Let us take, in the previous equality, the conditional expectation w.r.t $\mathcal{F}_n$ — denoted $\mathbb{E}_n$.

$$\begin{aligned}\mathbb{E}_n\left(\left\|\Delta_{n+\frac{1}{2}}\right\|^2\right) \le \|\Delta_n\|^2 &-2\gamma_n\Delta_n{}'(Q+y(X_n-x^\star))\Delta_n\\ &+ \gamma_{n+1}\mathbb{E}_n(\|\delta M_{n+1}\|^2) + \mathcal{O}\left(\frac{1}{n}\right)\|\Delta_n\|^2.\end{aligned}$$

Now, we can specify the definition the $\mathcal{A}_n$ sets a little more.

Since X_n converges almost surely to $x^\star$,

$$\forall \varepsilon > 0,\ \forall \mu > 0, \exists N > 0 \text{ such that } \forall n \geq N,\ \mathbb{P}\left(\sup_{m>n} \|X_m - x^\star\| > \mu\right) < \varepsilon. \tag{2.19}$$

Let $\lambda > 0$ be the smallest eigenvalue of A. Since A is symmetric definite positive, $\lambda > 0$. $\lim_{\|x\|\to 0} y(x) = 0$, so for $x < x_0$, $\|y(x)\| < \lambda/2$. Let $\varepsilon > 0$. Thanks to (2.19), there exists a rank N_0 — only depending on x_0 and ε — such that $\mathbb{P}(\sup_{m>N_0} \|X_m - x^\star\| > x_0) < \varepsilon$.

In the definition of the $\mathcal{A}_n$ sets (see (2.13)), we choose N_0 as defined above (and greater than n_0). On the set $\mathcal{A}_n$, $A + y(X_n - x^\star)$ is a definite positive matrix with smallest eigenvalue greater than $\lambda/2$. Therefore ${\Delta_n}'(A + y(X_n - x^\star))\Delta_n > \lambda/2\, \|\Delta_n\|^2$. We can assume that for $n > N_0$, $\quad \mathcal{O}(\frac{1}{n}) \leq \lambda/4\gamma_n$.

$$\begin{aligned}
\mathbb{E}\left(\left\|\Delta_{n+\frac{1}{2}}\right\|^2 \mathbf{1}_{\{\mathcal{A}_n\}}\right) - \mathbb{E}\left(\|\Delta_n\|^2 \mathbf{1}_{\{\mathcal{A}_n\}}\right) &\leq -\gamma_n \frac{\lambda}{4}\mathbb{E}\left(\|\Delta_n\|^2 \mathbf{1}_{\{\mathcal{A}_n\}}\right) + c\gamma_n, \\
\mathbb{E}\left(\left\|\Delta_{n+\frac{1}{2}}\right\|^2 \mathbf{1}_{\{\mathcal{A}_{n+1}\}}\right) - \mathbb{E}\left(\|\Delta_n\|^2 \mathbf{1}_{\{\mathcal{A}_n\}}\right) &\leq -\gamma_n \frac{\lambda}{4}\mathbb{E}\left(\|\Delta_n\|^2 \mathbf{1}_{\{\mathcal{A}_n\}}\right) + c\gamma_n, \qquad (2.20)
\end{aligned}$$

where c is a positive constant.

Now we would like to replace $\Delta_{n+\frac{1}{2}}$ by Δ_{n+1}.

$$\begin{aligned}
\|\Delta_{n+1}\|^2 &= \frac{\|X_0 - x^\star\|^2}{\gamma_{n+1}} \mathbf{1}_{\{p_{n+1}\neq 0\}} + \left\|\Delta_{n+\frac{1}{2}}\right\|^2 \mathbf{1}_{\{p_{n+1}=0\}}, \\
\|\Delta_{n+1}\|^2 &\leq \left\|\Delta_{n+\frac{1}{2}}\right\|^2 + \frac{\|X_0 - x^\star\|^2}{\gamma_{n+1}} \mathbf{1}_{\{X_n - \gamma_{n+1}U(X_n, Z_{n+1}) \notin \mathcal{K}_{\sigma_n}\}}.
\end{aligned}$$

Taking the conditional expectation w.r.t. $\mathcal{F}_n$ gives

$$\mathbb{E}_n \|\Delta_{n+1}\|^2 \leq \mathbb{E}_n \left\|\Delta_{n+\frac{1}{2}}\right\|^2 + \frac{\|X_0 - x^\star\|^2}{\gamma_{n+1}} \mathbb{E}_n\left(\mathbf{1}_{\{X_n - \gamma_{n+1}U(X_n, Z_{n+1}) \notin \mathcal{K}_{\sigma_n}\}}\right).$$

Multiplying by $\mathbf{1}_{\{\mathcal{A}_n\}}$ and noticing that $\mathcal{A}_{n+1} \subset \mathcal{A}_n$, we get

$$\begin{aligned}
\mathbb{E}\left(\|\Delta_{n+1}\|^2 \mathbf{1}_{\{\mathcal{A}_{n+1}\}}\right) \leq{}& \mathbb{E}\left(\left\|\Delta_{n+\frac{1}{2}}\right\|^2 \mathbf{1}_{\{\mathcal{A}_n\}}\right) \\
&+ \frac{\|X_0 - x^\star\|^2}{\gamma_{n+1}} \mathbb{E}\left(\mathbf{1}_{\{\mathcal{A}_n\}} \mathbb{E}_n\left(\mathbf{1}_{\{X_n - \gamma_{n+1}U(X_n, Z_{n+1}) \notin \mathcal{K}_{\sigma_n}\}}\right)\right). \qquad (2.21)
\end{aligned}$$

The conditional expectation on the right hand side can be rewritten

$$\begin{aligned}
\mathbb{E}_n\left(\mathbf{1}_{\{X_n - \gamma_{n+1}U(X_n, Z_{n+1}) \notin \mathcal{K}_{\sigma_n}\}} \mathbf{1}_{\{\mathcal{A}_n\}}\right) &\leq \mathbb{P}_n\left(\gamma_{n+1} \|U(X_n, Z_{n+1})\| \geq (X_n, \partial\mathcal{K}_{\sigma_n})\right) \mathbf{1}_{\{\mathcal{A}_n\}}, \\
&\leq \frac{\gamma_{n+1}^2}{d\left(X_n, \partial\mathcal{K}_{\sigma_n}\right)^2} \mathbb{E}_n\left(\|U(X_n, Z_{n+1})\|^2\right) \mathbf{1}_{\{\mathcal{A}_n\}}. \qquad (2.22)
\end{aligned}$$

Moreover, using the triangle inequality we have

$$d\left(X_n, \partial\mathcal{K}_{\sigma_n}\right) \geq d\left(x^\star, \partial\mathcal{K}_{\sigma_n}\right) - \|X_n - x^\star\|.$$

Using Hypothesis ($\mathcal{A}2.3$), $d\left(x^\star, \partial\mathcal{K}_{\sigma_n}\right) < \mu$ and on $\mathcal{A}_n$, $\quad \|X_n - x^\star\| \leq x_0$. Hence,

$$d\left(X_n, \partial\mathcal{K}_{\sigma_n}\right) \geq \mu - x_0.$$

One can choose x_0 smaller than $\mu/2$ for instance, so that $(\mu - x_0)^2 > \frac{\mu^2}{4}$. Thus, Equation (2.22) becomes

$$\mathbb{E}\left(\mathbf{1}_{\{X_n - \gamma_{n+1}U(X_n, Z_{n+1}) \notin \mathcal{K}_{\sigma_n}\}}\mathbf{1}_{\{\mathcal{A}_n\}}\right) \leq \frac{4\gamma_{n+1}^2}{\mu^2}\mathbb{E}\left(\|U(X_n, Z_{n+1})\|^2 \mathbf{1}_{\{\mathcal{A}_n\}}\right).$$

Thanks to Hypothesis ($\mathcal{A}2.2$) and the continuity of u, we get

$$\mathbb{E}\left(\|U(X_n, Z_{n+1})\|^2 \mathbf{1}_{\{\mathcal{A}_n\}}\right) \quad \leq \quad 2\sup_n \mathbb{E}(\|\delta M_{n+1}\|^2 \mathbf{1}_{\{\mathcal{A}_n\}}) + 2\sup_{\|X - x^\star\| < x_0} u(X)^2.$$

So, we get

$$\mathbb{E}\left(\mathbf{1}_{\{X_n - \gamma_{n+1}U(X_n, Z_{n+1}) \notin \mathcal{K}_{\sigma_n}\}}\mathbf{1}_{\{\mathcal{A}_n\}}\right) \leq c\gamma_{n+1}^2.$$

Hence, from Equation (2.21) we can deduce

$$\mathbb{E}\left(\|\Delta_{n+1}\|^2 \mathbf{1}_{\{\mathcal{A}_{n+1}\}}\right) \leq \mathbb{E}\left(\left\|\Delta_{n+\frac{1}{2}}\right\|^2 \mathbf{1}_{\{\mathcal{A}_n\}}\right) + c\gamma_n. \tag{2.23}$$

By combining Equations (2.23) and (2.20), we come up with

$$\mathbb{E}\left(\|\Delta_{n+1}\|^2 \mathbf{1}_{\{\mathcal{A}_{n+1}\}}\right) \quad \leq \quad \left(1 - \gamma_n \frac{\lambda}{4}\right)\mathbb{E}\left(\|\Delta_n\|^2 \mathbf{1}_{\{\mathcal{A}_n\}}\right) + c\gamma_n.$$

Let $\mathcal{I} = \left\{i > N_0 : -\frac{\lambda}{4}\mathbb{E}\left(\|\Delta_i\|^2 \mathbf{1}_{\{\mathcal{A}_i\}}\right) + c > 0\right\}$, then

$$\sup_{i \in \mathcal{I}} \mathbb{E}\left(\|\Delta_i\|^2 \mathbf{1}_{\{\mathcal{A}_i\}}\right) < \frac{4c}{\lambda} < \infty.$$

Otherwise for $i \notin \mathcal{I}$,

$$\mathbb{E}\left(\|\Delta_{i+1}\|^2 \mathbf{1}_{\{\mathcal{A}_{i+1}\}}\right) - \mathbb{E}\left(\|\Delta_i\|^2 \mathbf{1}_{\{\mathcal{A}_i\}}\right) \leq 0.$$

We will prove by recursion that $\forall i \geq N_0, \quad \mathbb{E}\left(\|\Delta_i\|^2 \mathbf{1}_{\{\mathcal{A}_i\}}\right) \leq \frac{4c}{\lambda} + \mathbb{E}\left(\|\Delta_{N_0}\|^2 \mathbf{1}_{\{\mathcal{A}_{N_0}\}}\right)$. It is obviously true for $i = N_0$. Let us assume that the recursion assumption holds for rank $i > N_0$. If $i+1 \in \mathcal{I}$, then $\mathbb{E}\left(\|\Delta_{i+1}\|^2 \mathbf{1}_{\{\mathcal{A}_{i+1}\}}\right) \leq \frac{24}{\lambda}$. Otherwise if $i+1 \notin \mathcal{I}$, $\mathbb{E}\left(\|\Delta_{i+1}\|^2 \mathbf{1}_{\{\mathcal{A}_{i+1}\}}\right) \leq \mathbb{E}\left(\|\Delta_i\|^2 \mathbf{1}_{\{\mathcal{A}_i\}}\right)$. So, using the hypothesis of recursion proves the result announced above. Therefore,

$$\sup_n \mathbb{E}\left(\|\Delta_n\|^2 \mathbf{1}_{\{\mathcal{A}_n\}}\right) < \infty.$$

In the end, this relation combined with (2.19) leads to the tightness of the sequence $(\Delta_n)_n$. Let $M > 0$.

$$\begin{aligned}\mathbb{P}(\|\Delta_n\| > M) &\le \mathbb{P}(\|\Delta_n\|(\mathbf{1}_{\{\mathcal{A}_n\}} + \mathbf{1}_{\{\mathcal{A}_n^c\}}) > M),\\ &\le \mathbb{P}(\|\Delta_n\|\,\mathbf{1}_{\{\mathcal{A}_n\}} > M/2) + \mathbb{P}(\|\Delta_n\|\,\mathbf{1}_{\{\mathcal{A}_n^c\}}) > M/2),\\ &\le 4/M^2\,\mathbb{E}\left(\|\Delta_n\|\,\mathbf{1}^2_{\{\mathcal{A}_n\}}\right) + \mathbb{P}(\mathcal{A}_n^c). \qquad (2.24)\end{aligned}$$

There exists a value of M depending on ε such that both terms on the right hand-side of (2.24) are bounded above by ε. This proves the tightness of $(\Delta_n)_n$ and ends to prove Lemma 2.3.1.

Proof of Lemma 2.3.3

If we go back to equation (2.18) and sum up this equality from n — chosen greater than N_0 — to $n+p$, we obtain

$$\begin{aligned}\Delta_{n+p} &= \Delta_n - \sum_{k=0}^{p-1}\gamma_{n+k}(\,Q\, + y(X_{n+k}-x^\star))\Delta_{n+k} + \sqrt{\gamma_{n+k+1}}\delta M_{n+k+1}\\ &\quad + \sum_{k=0}^{p-1}\sqrt{\gamma_{n+k+1}}p_{n+k+1} + \frac{\beta_{n+k}}{n+k+1}(B + y(X_{n+k}-x^\star))\Delta_{n+k}.\end{aligned}$$

We choose $u > 0$ such that $t_n(u) = p$. Since $X_n(\cdot)$ is piecewise constant on the subdivision defined by sequence $(\gamma_{n+p})_{p\ge 0}$, the discrete sums can be interpreted as integrals.

$$\Delta_n(u) = \Delta_n(0) - \int_0^u \left(Q + y(X_n(s)-x^\star)\right)\Delta_n(s)ds - W_n(u) + R_n(u) + P_n(u), \quad (2.25)$$

where

$$\begin{aligned}P_n(u) &= \sum_{k=0}^{t_n(u)}\sqrt{\gamma_{n+k+1}}p_{n+k+1},\\ R_n(u) &= \sum_{k=0}^{t_n(u)}\frac{\beta_{n+k}}{n+k+1}\,(B + y(X_{n+k}-x^\star))\Delta_{n+k}.\end{aligned}$$

Note that

$$\|R_n(u)\| \le \frac{C}{n^{1-\alpha}}\int_0^u (1 + \|y(X_n(s)-x^\star)\|)\,\|\Delta_n(s)\|\,ds.$$

Let $t > 0$ and $l = n + t_n(t)$. Note that on the set $\mathcal{A}_l$ $P_n(u) = 0$ $a.s.$ for all $u \le t$ and

$$\|y(X_n(s)-x^\star))\Delta_n(s)\|^2\,\mathbf{1}_{\{\mathcal{A}_l\}} \le \lambda/2\,\|\Delta_n(s)\|^2\,.$$

Using equation (2.25), we will show that $\left(\sup_{0\le t\le T} \|\Delta_n(t)\|^2\right)_n$ is tight in $\mathbb{R}$. Let us take the square and then the supremum over $[0,t]$ of Equation (2.25)

$$\begin{aligned}\sup_{u\le t} \|\Delta_n(u)\|^2 \mathbf{1}_{\{\mathcal{A}_l\}} &\le C' \|\Delta_n(0)\|^2 \mathbf{1}_{\{\mathcal{A}_l\}} + C'\, t \int_0^t \sup_{s\le u} \|\Delta_n(s)\|^2 \mathbf{1}_{\{\mathcal{A}_l\}} du \\ &\quad + C' \sup_{u\le t} \|W_n(u)\|^2 \mathbf{1}_{\{\mathcal{A}_l\}} + C' \sup_{u\le t} \|R_n(u)\|^2 \mathbf{1}_{\{\mathcal{A}_l\}}. \quad (2.26)\end{aligned}$$

The last term is bounded by $Ct \int_0^t \sup_{s\le u} \|\Delta_n(s)\|^2 \mathbf{1}_{\{\mathcal{A}_l\}} du$.
We define $e_n(t) = \mathbb{E}\left(\sup_{u\le t} \|\Delta_n(u)\|^2 \mathbf{1}_{\{\mathcal{A}_l\}}\right)$, then taking expectation in (2.26) gives

$$e_n(t) \le C e_n(0) + Ct \int_0^t e_n(u) du + C\, \mathbb{E}\left(\sup_{u\le t} \|W_n(u)\|^2 \mathbf{1}_{\{\mathcal{A}_l\}}\right). \quad (2.27)$$

Doob's inequality applied to $(W_n(u)\mathbf{1}_{\{\mathcal{A}_{n+t_n(u)}\}})_{0\le u\le t}$ enables us to rewrite (2.27)

$$e_n(t) \le C e_n(0) + Ct \int_0^t e_n(u) du + 4C\, \mathbb{E}\left\|W_n(t)\mathbf{1}_{\{\mathcal{A}_l\}}\right\|^2.$$

Thanks to Lemma 2.3.1, $\sup_n e_n(0) < \infty$. Hence, $e_n(0)$ can be incorporated into constant C, which remains independent of n.
$\mathbb{E}\left\|W_n(t)\mathbf{1}_{\{\mathcal{A}_l\}}\right\|^2 \le \sum_{i=n+1}^{n+t_n(t)+1} \gamma_i \mathbb{E}(\|\delta M_i\|^2 \mathbf{1}_{\{\|X_{i-1}-x^\star\|\le\eta\}})$. So, $\sup_n \mathbb{E}\left\|W_n(t)\mathbf{1}_{\{\mathcal{A}_{n+t_{n(t)}}\}}\right\|^2$ is bounded by $\kappa_0 t$. Then, we come up with the following inequality for any $n > N_0$

$$e_n(t) \le C(1+t) + CT \int_0^t e_n(u) du, \text{ for all } t \text{ in } [0,T],$$

where constant C depends neither on n nor on T.
Using Bellman-Gronwall's inequality, we obtain a key upper-bound for $e_n(t)$

$$e_n(t) \le C(1+t)e^{CT^2}, \text{ for all } t \text{ in } [0,T] \text{ and } n > N_0.$$

The previous inequality can be summed up as

$$\sup_n \mathbb{E}\left(\sup_{t\le T} \|\Delta_n(t)\|^2 \mathbf{1}_{\{\mathcal{A}_{n+t_n(T)}\}}\right) < \infty \text{ for any } T. \quad (2.28)$$

Equation (2.28) implies that $\left(\sup_{t\le T} \|\Delta_n(t)\|^2\right)_n$ is tight in $\mathbb{R}$.
From now on we define $\bar{e} = \sup_n \mathbb{E}\left(\sup_{t\le T} \|\Delta_n(t)\|^2 \mathbf{1}_{\{\mathcal{A}_{n+t_n(T)}\}}\right)$.

Proof of Lemma 2.3.4

Proving Lemma 2.3.4 straightforwardly is pretty hard and instead we consider $\widetilde{W}$

$$\widetilde{W}_n(0) = 0 \quad \text{and} \quad \widetilde{W}_n(t) = \sum_{i=n+1}^{n+t_n(t)+1} \sqrt{\gamma_i}\delta M_i \mathbf{1}_{\{\mathcal{A}_{i-1}\}} \quad \text{for } t > 0. \tag{2.29}$$

Remark 2.3.7. *We consider N_0 and ε as defined by Equation (2.41). Since for all $n > N$, and for all $a > 0$, $\mathbb{P}\left(\sup_{t\leq T}\left\|W_{n(t)} - \widetilde{W}_n(t)\right\| > a\right) \leq \varepsilon$ for any $T > 0$, it is sufficient to prove Lemma 2.3.4 for $\widetilde{W}_n$ and it will automatically hold for W_n.*

First, we prove that $(\widetilde{W}_n(\cdot))_n$ is $\mathcal{C}-$tight and that $(\sup_{t\in[0,T]}\left\|\widetilde{W}_n(t)\right\|)_n$ is tight in $\mathbb{R}$. These two points imply that $(\widetilde{W}_n(\cdot))_n$ is tight in $\mathbb{D}[0,T]$ and that every converging subsequence converges in law to a continuous process. Then, we prove that any such limit is a martingale. Finally, we establish that these limit martingales have predictable quadratic variation equal to Σt. Thanks to Lévy's Theorem[2], combining these two points imply that W is a Wiener process with covariance matrix Σ.

Tightness of $(\widetilde{W}_n(\cdot))_n$ in $\mathbb{D}[0,T]$. We have already seen that $\sup_n E\left(\left\|\widetilde{W}_n(t)\right\|^2\right) \leq \kappa_0 t$, so the family $\{\widetilde{W}_n(t); n \geq 1\}$ is uniformly integrable for each t in $[0,T]$.

Moreover using Doob's inequality, it comes that

$$E\left(\sup_{t\in[0,T]}\left\|\widetilde{W}_n(t)\right\|^2\right) \leq 4\sum_{i=n+1}^{n+1+t_n(T)} \gamma_i\, \mathbb{E}\left(\|\delta M_i\|^2 \mathbf{1}_{\{\|X_{i-1}-x^\star\|\leq\eta\}}\right) \leq 4\kappa_0 T.$$

Thus, $\left(\sup_{t\in[0,T]}\left\|\widetilde{W}_n(t)\right\|\right)_n$ is tight in $\mathbb{R}$.

We want to prove that the sequence of processes $(\widetilde{W}_n(t))_{0\leq t\leq T}$ satisfies a $\mathcal{C}-$tightness criterion. It suffices to show that there exist two positive real numbers λ and β such that for any (t,s) in $[0,T]^2$ the following inequality holds

$$\mathbb{E}(\left\|\widetilde{W}_n(t) - \widetilde{W}_n(s)\right\|^\lambda) \leq \kappa\,|t-s|^{1+\beta}.$$

Let us choose a couple (s,t) in $[0,T]^2$ such as $s < t$ and an $\lambda > 0$. Using Burkholder-

[2] see Protter [52, p. 86] for instance.

Davis-Gundy's inequality[3] we can write

$$\begin{aligned}
\mathbb{E}\left(\left\|\widetilde{W}_n(t)-\widetilde{W}_n(s)\right\|^{\lambda}\right) &= \mathbb{E}\left(\left\|\widetilde{W}_{n+t_n(t)}-\widetilde{W}_{n+t_n(s)}\right\|^{\lambda}\right), \\
&\leq \kappa\,\mathbb{E}\left\|\left[\widetilde{W}_{n+t_n(.)}-\widetilde{W}_{n+t_n(s)}\right]_t\right\|^{\lambda/2}, \\
&\leq \kappa\,\mathbb{E}\left\|\sum_{i=n+1+t_n(s)}^{n+1+t_n(t)} \gamma_i\,\delta M_i \delta M_i' \mathbf{1}_{\{\mathcal{A}_{i-1}\}}\right\|^{\lambda/2}, \qquad (2.30)
\end{aligned}$$

Now, we use a well-known inequality for convex functions, assuming that $\lambda > 2$.

$$\begin{aligned}
\left\|\sum_{i=n+1+t_n(s)}^{n+1+t_n(t)} \gamma_i\,\delta M_i \delta M_i' \mathbf{1}_{\{\mathcal{A}_{i-1}\}}\right\|^{\lambda/2} &\leq |t-s|^{\lambda/2}\left\|\sum_{i=n+1+t_n(s)}^{n+1+t_n(t)} \frac{\gamma_i}{t-s}\,\|\delta M_i\|^2\,\mathbf{1}_{\{\mathcal{A}_{i-1}\}}\right\|^{\lambda/2}, \\
&\leq |t-s|^{\lambda/2-1}\sum_{i=n+1+t_n(s)}^{n+1+t_n(t)} \gamma_i\,\|\delta M_i\|^{\lambda}\,\mathbf{1}_{\{\mathcal{A}_{i-1}\}}.
\end{aligned}$$

Thus, the expectation on the right hand side of (2.30) is bounded by

$$|t-s|^{\lambda/2-1}\sum_{i=n+1+t_n(s)}^{n+1+t_n(t)} \gamma_i\,\mathbb{E}(\|\delta M_i\|^{\lambda}\,\mathbf{1}_{\{\mathcal{A}_{i-1}\}}).$$

We choose $\lambda = 2+\rho$ — ρ being defined in Hypothesis ($\mathcal{A}2.2$) — to obtain the desired inequality

$$\mathbb{E}\left(\left\|\widetilde{W}_n(t)-\widetilde{W}_n(s)\right\|^{\lambda}\right) \leq \kappa\ |t-s|^{\rho/2+1}.$$

The tightness of $(\widetilde{W}_n(0)_n$ is given by its uniform square integrability. Thus, the sequence of processes $(\widetilde{W}_n(\cdot))_n$ is $\mathcal{C}$-tight. Moreover, thanks to Lemma 2.3.3 $(\sup_{t\in[0,T]}\left\|\widetilde{W}_n(t)\right\|)_n$ is tight in $\mathbb{R}$. Hence, $(\widetilde{W}_n(\cdot))_n$ is tight in $\mathbb{D}$.

Any converging subsequence converges in law to a continuous martingale. Let $(\widetilde{W}_n(\cdot))_n$ denote a converging subsequence with limit W. We will show that W is a continuous martingale. Since $(\widetilde{W}_n(\cdot))_n$ is $\mathcal{C}-$tight, W is a continuous process.

For any $L > 0$, we define the continuous function f_L such that $f_L(x) = x$ if $\mathbf{1}_{\{\|x\|\leq L\}}$ and $f_L(x) = 0$ if $\mathbf{1}_{\{\|x\|\geq L+1\}}$. Therefore, f_L is a continuous bounded function. We have for all $n > 0$

$$\kappa_0 t \geq \mathbb{E}(\left\|\widetilde{W}_n(t)\right\|^2) \geq \mathbb{E}(f_L(\left\|\widetilde{W}_n(t)\right\|^2)).$$

[3] see Protter [52, Theorem 48] for càdlàg martingales.

Thanks to the convergence of $(\widetilde{W}_n(\cdot))_n$, we get[4]

$$\kappa_0 t \geq \mathbb{E}(f_L(\left\|\widetilde{W}(t)\right\|^2)).$$

$f_L(\|W(t)\|^2)$ is non decreasing w.r.t to L and positive so using the monotone convergence Theorem, we obtain

$$\kappa_0 t \geq \mathbb{E}(\|W(t)\|^2). \tag{2.31}$$

This proves that W is square integrable.

Let h be a continuous bounded function on $\mathbb{D}$. Since $\widetilde{W}_n(\cdot)$ converges in law in $\mathbb{D}$ to W, we have for all $0 < s < t \leq T$

$$\mathbb{E}\left[h(\widetilde{W}_n(u); u \leq s) f_L(\widetilde{W}_n(t) - \widetilde{W}_n(s))\right] \xrightarrow[n\to\infty]{} \mathbb{E}\left[h(W(u); u \leq s) f_L(W(t) - W(s))\right]. \tag{2.32}$$

$\lim_{L\to\infty} f_L(W(t) - W(s)) = W(t) - W(s)$. Thanks to (2.31), we can use the bounded convergence Theorem to show that the expectation on the right hand side of (2.32) tends to $\mathbb{E}\left[h(W(u); u \leq s)(W(t) - W(s))\right]$ when L goes to infinity.

Thanks to the uniform integrability of $(\widetilde{W}_n(t))_n$ for each fixed t,

$$\sup_n \mathbb{E}\left[h(\widetilde{W}_n(u); u \leq s)\left\{f_L(\widetilde{W}_n(t) - \widetilde{W}_n(s)) - (\widetilde{W}_n(t) - \widetilde{W}_n(s))\right\}\right] \xrightarrow[L\to\infty]{} 0.$$

So,

$$\mathbb{E}\left[h(\widetilde{W}_n(u); u \leq s)(\widetilde{W}_n(t) - \widetilde{W}_n(s))\right] \xrightarrow[n\to\infty]{} \mathbb{E}\left[h(W(u); u \leq s)(W(t) - W(s))\right]. \tag{2.33}$$

Since for any fixed n $(\widetilde{W}_n(t))_t$ is a martingale, $\mathbb{E}\left[h(\widetilde{W}_n(u); u \leq s)(\widetilde{W}_n(t) - \widetilde{W}_n(s))\right] = 0$. Then, we come up with $\mathbb{E}\left(h(W(u); u \leq s)(W(t) - W(s))\right) = 0$, which proves that W is a martingale w.r.t. the filtration it generates.

Any limit W has predictable quadratic variation equal to Σt. Since the predictable quadratic variation process is unique, up to an evanescent set, it is sufficient to prove that $(W_t W_t' - \Sigma t)_t$ is a martingale.

As $\mathbb{E}\left\|\widetilde{W}_n(t)\right\|^{2+\rho}$ is uniformly bounded in n, $(\widetilde{W}_n(t)\widetilde{W}_n(t)')_n$ is uniformly integrable for each fixed t. So using truncation functions as above, it is straightforward to prove that W is square integrable. Moreover,

[4]In fact, we also need the continuity of $\omega \in \mathbb{D} \longmapsto \omega(t)$ on a set of measure 1 for the law of W. $\omega \in \mathbb{D} \longmapsto \omega(t)$ is continuous for the topology on $\mathbb{D}$ at every point X such that X does not jump at time t. Therefore, the coordinate applications are continuous on the set of continuous paths which is of measure 1 for the law of W because W is a.s. continuous. Hence, $\omega \in \mathbb{D} \longmapsto f_L(\|\omega(t)\|)$ is continuous for the topology on $\mathbb{D}$ on a set of measure 1 for the limiting law.

$$\begin{aligned}
\mathbb{E}\left(\left\|\langle \widetilde{W}_n, \widetilde{W}_n\rangle_t\right\|^{1+\rho/2}\right) &\leq \mathbb{E}\left(\left(\sum_{i=n+1}^{n+1+t_n(t)} \gamma_i \mathbb{E}(\|\delta M_{i+1}\|^2 \mathbf{1}_{\{\mathcal{A}_{i-1}\}}|\mathcal{F}_i)\right)^{1+\rho/2}\right), \\
&\quad \text{using a convexity inequality, we get} \\
&\leq t^{\rho/2} \sum_{i=n+1}^{n+1+t_n(t)} \gamma_i \mathbb{E}(\|\delta M_{i+1}\|^{2+\rho} \mathbf{1}_{\{\mathcal{A}_{i-1}\}}), \\
&\leq t^{1+\rho/2}\kappa.
\end{aligned}$$

So, $(\langle \widetilde{W}_n, \widetilde{W}_n\rangle_t)_n$ is uniformly integrable for each t.

Let h be a continuous bounded function on $\mathbb{D}$. Using Hypothesis ($\mathcal{A}$2.2), we can see that $(\langle \widetilde{W}_n, \widetilde{W}_n\rangle_t)$ tends in probability to Σt and thanks to the uniform integrability, the convergence also occurs in $\mathbb{L}^1$. Hence,

$$\lim_n \mathbb{E}\left[h(\widetilde{W}_n(u); u \leq s)(\langle \widetilde{W}_n, \widetilde{W}_n\rangle_t - \Sigma t)\right] = 0. \tag{2.34}$$

Since $\widetilde{W}_n(t)$ is martingale for any fixed n,

$$\mathbb{E}\left[h(\widetilde{W}_n(u); u \leq s)(\widetilde{W}_n(t)\widetilde{W}_n(t)' - \langle \widetilde{W}_n, \widetilde{W}_n\rangle_t)\right] = \\ \mathbb{E}\left[h(\widetilde{W}_n(u); u \leq s)(\widetilde{W}_n(s)\widetilde{W}_n(s)' - \langle \widetilde{W}_n, \widetilde{W}_n\rangle_s)\right].$$

Once again, we use truncation functions. Since $(\widetilde{W}_n(\cdot))_n$ converges in law in $\mathbb{D}$ to W and $(\widetilde{W}_n(t)\widetilde{W}_n(t)')_n$ is uniformly integrable for each t, $\mathbb{E}\left[h(\widetilde{W}_n(u); u \leq s)(\widetilde{W}_n(t)\widetilde{W}_n(t)' - \Sigma t)\right]$ tends to $\mathbb{E}\left[h(W(u); u \leq s)(W(t)W(t)' - \Sigma t)\right]$.

Consequently using Equation (2.34), we get

$$\mathbb{E}\left[h(W(u); u \leq s)(W(t)W(t)' - \Sigma t)\right] = \mathbb{E}\left[h(W(u); u \leq s)(W(s)W(s)' - \Sigma s)\right].$$

Thus, $(W(t)W(t)' - \Sigma t)_t$ is a martingale. Since the predictable quadratic variation process is unique, up to an evanescent set, $\langle W, W\rangle_t = \Sigma t$ a.s.. Moreover W is continuous, so Lévy's characterisation of the Wiener process proves that W is a Wiener process with covariance matrix Σ.

Hence, any converging subsequence of $(\widetilde{W}_n(\cdot))_n$ converge to a Wiener process with covariance matrix Σ, which implies that the whole sequence converges in law to the process W. Thanks to Remark 2.3.7, we know that W is also the limit of $W_n(\cdot)$.

Proof of Lemma 2.3.5

Let us choose some fixed positive μ, ε and a corresponding δ. S and τ stands for two stopping times as introduced in Lemma 2.3.5. Let $l = n + t_n(T)$.

$$\mathbb{P}\left(\|\Delta_n(\tau) - \Delta_n(S)\| \geq 2\mu\right) \leq \mathbb{P}\left(\|\Delta_n(\tau) - \Delta_n(S)\| \mathbf{1}_{\{\mathcal{A}_l\}} \geq \mu\right) + \mathbb{P}(\mathcal{A}_l^c).$$

Remember that $\mathbb{P}(\mathcal{A}_l^c) \leq \varepsilon$.

$$\mathbb{P}\left(\|\Delta_n(\tau)-\Delta_n(S)\|\,\mathbf{1}_{\{\mathcal{A}_l\}} \geq \mu\right) \leq \mathbb{P}\left(\left\|\int_S^\tau \left(Q - y(X_n(u)-x^\star)\right)\Delta_n(u)\mathbf{1}_{\{\mathcal{A}_l\}}du\right\| \geq \frac{\mu}{6}\right)$$
$$+\ \mathbb{P}\left(\|W_n(\tau)-W_n(S)\|\,\mathbf{1}_{\{\mathcal{A}_l\}} \geq \frac{\mu}{6}\right) + \mathbb{P}\left(\|R_n(\tau)-R_n(S)\|\,\mathbf{1}_{\{\mathcal{A}_l\}} \geq \frac{\mu}{6}\right). \quad (2.35)$$

On the set $\mathcal{A}_l$, $P_n(u) = 0$ a.s. for all $u \leq T$.

The first term is handled using Markov's inequality

$$\begin{aligned}
&\mathbb{P}\left(\left\|\int_S^\tau \left(Q - y(X_n(u)-x^\star)\right)\Delta_n(u)du\mathbf{1}_{\{\mathcal{A}_n\}}\right\| \geq \frac{\mu}{6}\right)\\
&\qquad \leq \frac{c}{\mu^2}\mathbb{E}\left(\delta\int_S^{S+\delta}\left\|\Delta_n(u)\mathbf{1}_{\{\mathcal{A}_l\}}\right\|^2 du\right),\\
&\qquad \leq \frac{c}{\mu^2}\mathbb{E}\left(\delta\int_0^{T}\left\|\Delta_n(u)\mathbf{1}_{\{\mathcal{A}_l\}}\right\|^2 du\right),\\
&\qquad \leq \frac{c\delta}{\mu^2}\bar{e}\,T,\\
&\qquad \leq \frac{c}{\mu^2}K, \text{ where } c \text{ is a positive constant only depending on } T.
\end{aligned}$$

The third term can be treated like the first one. Now, we will apply Burkholder-Davis-Gundy's inequality to the stopped martingale $((W_n(t)-W_n(t\wedge S))\mathbf{1}_{\{\mathcal{A}_{n+t_n(t)}\}})_t$.

$$\begin{aligned}
\mathbb{E}(\|W_n(\tau)-W_n(S)\|^{2+\rho}\,\mathbf{1}_{\{\mathcal{A}_{n+t_n(t)}\}}) &\leq \mathbb{E}\left(\sum_{i=n+1+t_n(S)}^{n+1+t_n(S+\delta)}\gamma_i\ \|\delta M_i\|^2\right)\mathbf{1}_{\{\mathcal{A}_{i-1}\}}^{1+\rho/2},\\
&\text{using a convexity inequality, we obtain}\\
&\leq \delta^{1+\rho/2}\mathbb{E}\left(\sum_{i=n+1+t_n(S)}^{n+1+t_n(S+\delta)}\frac{\gamma_i}{\delta}\ \|\delta M_i\|^{2+\rho}\,\mathbf{1}_{\{\mathcal{A}_{i-1}\}}\right),\\
&\leq \delta^{\rho/2}\left(\sum_{i=n+1}^{n+1+t_n(T)}\gamma_i\,\mathbb{E}(\|\delta M_i\|^{2+\rho}\,\mathbf{1}_{\{\mathcal{A}_{i-1}\}})\right).
\end{aligned}$$

Using hypothesis ($\mathcal{A}2.2$), we come up with the following upper-bound

$$\mathbb{E}\,\|W_n(\tau)-W_n(S)\|^{2+\rho} \leq \delta^{\rho/2}T\sup_i \mathbb{E}(\|\delta M_i\|^{2+\rho}\,\mathbf{1}_{\{\mathcal{A}_{i-1}\}}).$$

Finally, we obtain a new upper bound in (2.35)

$$\mathbb{P}\left(\|\Delta_n(\tau)-\Delta_n(S)\| \geq \mu\right) \leq \delta^{\rho/2}\left(\frac{C_1}{\mu^2}+\frac{C_2}{\mu^{2+\rho}}\right)+\varepsilon, \quad (2.36)$$

where C_1 and C_2 are two positive constants independent of S, τ, n and μ. Assuming that $\mu < 1$, Equation (2.36) becomes

$$\mathbb{P}\left(\|\Delta_n(\tau) - \Delta_n(S)\| \geq \mu\right) \leq \delta^{\rho/2} \frac{C}{\mu^{2+\rho}} + \varepsilon, \tag{2.37}$$

where C is a positive constant.

Choosing $\delta = (\varepsilon\mu^{2+\rho})^{1/\rho}$ shows that property (2.15) holds true. Since $\left(\sup_{t\in[0,T]} \|\Delta_n(t)\|\right)_n$ is tight, Equation (2.37) ends to prove that $(\Delta_n(\cdot))_n$ is tight in $\mathbb{D}$.

Proof of Lemma 2.3.6

$(W_n(\cdot))_n$ is $C-$tight and $(\Delta_n(\cdot))_n$ is tight, so it is quite straightforward[5] that the couple $(W_n(\cdot), \Delta_n(\cdot))_n$ is tight in $\mathbb{D} \times \mathbb{D}$. For a proof of the result, one can see Jacod and Shiryaev [40, Corollary 3.33, page 317].

Thus, we can extract a converging subsequence $(W_{\phi(n)}(\cdot), \Delta_{\phi(n)}(\cdot))$ with limit $(W^\phi(\cdot), \Delta^\phi(\cdot))$. We will prove that in Equation (2.25), $\left(\sup_{0\leq u\leq T} \left\|R_{\phi(n)}(u)\right\|\right)_n$ and $\left(\sup_{0\leq u\leq T} \left\|P_{\phi(n)}(u)\right\|\right)_n$ tend to zero in probability.

Let $\mu > 0$ and $l = \phi(n) + t_{\phi(n)}(T)$.

$$\mathbb{P}\left(\sup_{0\leq u\leq T} \left\|R_{\phi(n)}(u)\right\| > \mu\right) \leq \mathbb{P}\left(\frac{C}{\phi(n)} \int_0^T \left(1 + \left\|y(X_{\phi(n)}(u) - x^\star)\right\|\right) \left\|\Delta_{\phi(n)}(u)\right\| du > \mu\right).$$

We split the probability on the right hand side on the sets $\mathcal{A}_l$ and $\mathcal{A}_l^c$.

$$\mathbb{P}\left(\frac{C}{\phi(n)} \int_0^T \left(1 + \left\|y(X_{\phi(n)}(u) - x^\star)\right\|\right) \left\|\Delta_{\phi(n)}(u)\right\| \mathbf{1}_{\{\mathcal{A}_l^c\}} du > \frac{\mu}{2}\right) \leq \mathbb{P}(\mathbf{1}_{\{\mathcal{A}_l^c\}}).$$

Now, we tackle the probability on $\mathcal{A}_l$.

$$\begin{aligned}
&\mathbb{P}\left(\frac{C}{\phi(n)} \int_0^T \left(1 + \left\|y(X_{\phi(n)}(u) - x^\star)\right\|\right) \left\|\Delta_{\phi(n)}(u)\right\| \mathbf{1}_{\{\mathcal{A}_l\}} du > \frac{\mu}{2}\right) \\
&\qquad\leq \frac{c}{\mu^2\phi(n)^2} \mathbb{E}\left(\int_0^T \left(1 + \left\|y(X_{\phi(n)}(u) - x^\star)\right\|\right) \left\|\Delta_{\phi(n)}(u)\right\| \mathbf{1}_{\{\mathcal{A}_l\}} du\right), \\
&\qquad\leq \frac{c}{\mu^2\phi(n)^2} \mathbb{E}\left(\int_0^T \left\|\Delta_{\phi(n)}(u)\right\| \mathbf{1}_{\{\mathcal{A}_l\}} du\right), \\
&\qquad\leq \frac{c}{\mu^2\phi(n)^2} T\bar{e}.
\end{aligned} \tag{2.38}$$

For n large enough, the term on the right hand side of (2.38) can be made smaller than ε.

[5]The pseudo continuity modulus, w', on $\mathbb{D}$ has no linearity property, but for any $\alpha, \beta \in \mathbb{D}$, any $\delta > 0$ we have $w'(\alpha + \beta, \delta) \leq w'(\alpha, \delta) + w(\beta, 2\delta)$, where w is the continuity modulus.

Then, it is then clear that for n large enough

$$\mathbb{P}\left(\sup_{0\leq u\leq T}\left\|R_{\phi(n)}(u)\right\|>\mu\right)\leq 2\varepsilon.$$

The term $\mathbb{P}\left(\sup_{0\leq u\leq T}\left\|P_{\phi(n)}(u)\right\|>\mu\right)$ can also be treated by splitting the probability on $\mathcal{A}_l$ and its complementary set. Recall that on the set $\mathcal{A}_l$, $P_n(u)=0$ a.s. for all $u\leq T$. Hence P_n and R_n both converge to the zero process in $\mathbb{D}$.

Remember that the integral is a continuous application from $\mathbb{D}$ into $\mathbb{R}$. More precisely for any real numbers a and b in $[0,T]$, the application $\omega\in\mathbb{D}\longmapsto\int_a^b\omega(t)dt$ is continuous.

Thanks to Lemma 2.3.4, $W^{\phi}(\cdot)$ is a Wiener process with covariance matrix Σ. Hence, the limit of $W_{\phi(n)}(\cdot)$ is independent of ϕ. So, letting n go to infinity in (2.25) enables to show that the limit $\Delta^{\phi}(\cdot)$ satisfies the following equation

$$\Delta^{\phi}(t)=\Delta^{\phi}(0)-\int_0^t Q\Delta^{\phi}(u)du-W(t), \tag{2.39}$$

which is equivalent to

$$d\Delta^{\phi}(t)=-Q\Delta^{\phi}(t)dt-dW(t).$$

Equation (2.39) shows that the set of all possible limits of any converging subsequence of $(\Delta_n(\cdot))_n$ is a family of Ornstein Uhlenbeck processes indexed by their initial conditions. So, if we manage to prove that the set $\{\Delta^{\phi}(0);\phi \text{ such that } \Delta_{\phi(n)}(\cdot) \text{ converges}\}$ is reduced to a single point, we will have stated the convergence of the whole sequence $(\Delta_n(\cdot))_n$ and not only of a subsequence. Any limit $\Delta(\cdot)$ satisfies

$$\Delta(t)=e^{-Qt}\Delta(0)-\int_0^t e^{Q(u-t)}dW(u).$$

The stochastic integral converges in distribution to a random normal variable with mean 0 and covariance matrix $\int_0^\infty e^{-Qu}\Sigma e^{-Qu}du$ as t goes to infinity. So does the process Δ since the set of all possible laws for $\Delta(0)$ is tight and e^{-Qt} tends to zero when t goes to infinity. This limit happens to be the unique stationary law for the Δ process.

Now, we want to prove that the set of all possible laws for $\Delta(0)$ is reduced to the stationary law described above. The way we prove it is widely inspired from Benveniste et al. [13] and Bouton [18].

Stationarity of any limit. Let $\boldsymbol{\nu}=\{\text{possible laws for }\Delta(0)\}$. $\boldsymbol{\nu}$ is a weakly compact set. For any $\nu\in\boldsymbol{\nu}$, let $P_\nu(t)$ denote the law at time t of the process $\Delta(\cdot)$ with initial law ν. Let f be a continuous bounded function on $\mathbb{R}^d$ and ν_g be the stationary law described above.

Let us choose an $\varepsilon>0$. Since $\boldsymbol{\nu}$ is weakly compact, there exists $T>0$ such that

$$|\langle f,P_\nu(t)\rangle-\langle f,\nu_g\rangle|\leq\varepsilon \tag{2.40}$$

for any $t>T$ and any $\nu\in\boldsymbol{\nu}$.

We fix such a $T > 0$ and choose $\nu \in \boldsymbol{\nu}$. We can extract a converging subsequence $(\Delta_{\phi(n)}(\cdot))_n$ such that ν is the initial laws of the limit.

We define

$$\psi(n) = \inf\left\{ k \geq 0; \sum_{i=k}^{\phi(n)} \gamma_i \leq T \right\}. \tag{2.41}$$

For n large enough, $\psi(n) > 0$ which means that $\psi(n) + t_{\psi(n)}(T) = \phi(n)$ and ψ is an increasing function. Hence we have the equality $\Delta_{\phi(n)}(0) = \Delta_{\psi(n)}(T)$. We can extract one more subsequence such that $\Delta_{\psi(\psi'(n))}(\cdot)$ converges. If ν' denotes the initial law of the limit, we have

$$|\langle f, \nu\rangle - \langle f, \nu_g\rangle| = |\langle f, P_{\nu'}(T)\rangle - \langle f, \nu_g\rangle| \leq \varepsilon.$$

The last part of the inequality comes from (2.40). This proves that $\nu = \nu_g$. Henceforth, any converging subsequence of $\Delta_n(\cdot)$ converges to a stationary Ornstein Uhlenbeck process.

W is a $\mathcal{F}^{\Delta,W}$−martingale. The only remaining point to prove is that $\Delta(0)$ is independent of $\sigma(W(t); t > 0)$. This is the same as proving that W is a $\mathcal{F}^{\Delta,W}$−Wiener process, where $\mathcal{F}_t^{\Delta,W}$ is the smallest σ−algebra that measures $\{\Delta(s), W(s); s \leq t\}$.

Since we already know that W is continuous and that $\langle W, W\rangle_t = t$ a.s., it is sufficient to prove that W is a $\mathcal{F}^{\Delta,W}$−martingale.

Let h be a continuous bounded function on $\mathbb{D}$. Since $(W_n, \Delta_n) \Longrightarrow (W, \Delta)$ and thanks to Remark 2.3.7, $(\widetilde{W}_n, \Delta_n) \Longrightarrow (W, \Delta)$. Moreover $(\widetilde{W}_n(t))_n$ is uniformly integrable for each t, it is quite obvious that

$$\mathbb{E}\left(h((\Delta_n(s), \widetilde{W}_n(s); s \leq t)(\widetilde{W}_n(t+\tau) - \widetilde{W}_n(t))\right) \xrightarrow[n\to\infty]{} \mathbb{E}\left(h((\Delta(s), W(s); s \leq t))(W(t+\tau) - W(t))\right).$$

$\widetilde{W}_n(\cdot)$ is a $\mathcal{F}_{n+t_n(\cdot)}$ martingale and $\Delta_n(\cdot)$ is measurable with respect to the shifted filtration. Hence,

$$\mathbb{E}\left(h((\Delta_n(s), \widetilde{W}_n(s); s \leq t)(\widetilde{W}_n(t+\tau) - \widetilde{W}_n(t))\right) = 0.$$

Consequently,

$$\mathbb{E}\left(h((\Delta(s), W(s); s \leq t))(W(t+\tau) - W(t))\right) = 0.$$

This last equality implies that W is a $\mathcal{F}^{\Delta,W}$−Wiener process.

Chapter 3

Practical applications : Calibration and Variance reduction

In this chapter, we present two examples of applications of stochastic algorithms. The first one deals with the complex problem of calibrating the correlation in a multi asset model in finance. To solve this problem, we propose to use a constrained stochastic algorithm. The second example we consider hereafter is based on the work of Arouna [7]. He describes an original way of implementing a variance reduction technique based on importance sampling for Monte Carlo computations. He proposes to use a stochastic approximation to compute the optimal change of probability (amongst a certain class of changes). We develop this example and compare his approach with the use of an averaging stochastic approximation as the one studied in Chapter 2. Meanwhile, we pragmatically improve his almost sure convergence result by relaxing the hypotheses on the increasing sequence of compact sets. We apply the technique to the computation of the price of Mountain options.

3.1 Calibration of a multi dimensional model

One of the major prises for banks is to find the best possible fitting model for a given market. This problem can be summarised as follows : for a fixed model depending on some parameters, the idea is to determine the set of parameters that matches best the market prices for liquid options. We are especially interested in finding the correlation parameter in multidimensional market models. Finding this parameter is definitely essential to settle a replicating strategy. So far, methods from the deterministic optimisation are used but their efficiency decreases as soon as the number of assets involved increases. In the following section, we present the mathematical modelling of the problem and explain how stochastic algorithms and especially the Robbins Monro algorithm can help solving such problems. Then, we handle the case of a basket option and show numerical results. We especially focus on the benefit of using averaging stochastic approximations.

3.1.1 Mathematical modelling of the problem

We consider a financial market modelled by a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. We assume that $\mathbb{P}$ is already the risk neutral measure and that $\mathcal{F} = (\mathcal{F}_t)_{t\geq 0}$ is the filtration generated by d one-dimensional standard Brownian motions $B^i = \{B^i_t, t \geq 0\}$. Under these conditions, the prices $\bar{S}^i_t$ of the assets are given by the following equation

$$d\bar{S}^i_t = \bar{S}^i_t(rdt + \sigma_i dB^i_t), \quad \bar{S}^i_0 = x_i,$$

where $d\langle B^i, B^j\rangle_t = \rho_{ij}dt \quad \text{if } i \neq j \quad$ and 1 otherwise, r is the interest rate and σ_i the volatility of asset i. Let $\Gamma = (\rho_{ij})_{1\leq i,j\leq d}$ be the correlation matrix defined above. Γ is the covariance matrix of the Brownian motions and must therefore be positive. We will also assume that Γ is definite.

We suppose to have for each asset some option prices given by the market, so that we are able to fit the parameters that describe each asset independently. Nevertheless, we may need to consider several assets together, because there are involved in the same basket option for instance. In this case, we need to calibrate the correlation parameter between the different assets. We assume to have some basket option prices denoted by $(C_j)_{j\in\mathcal{J}}$.

Our purpose is then to find the correlations which have lead to these prices.

We consider some payoffs $(\phi_j(\bar{S}_T))_{j\in\mathcal{J}}$ and their corresponding option prices $(C_j)_{j\in\mathcal{J}}$. The idea is to minimise, over the set of admissible matrices Γ, a well chosen criteria. We formulate the problem as a least square problem by considering the relative difference between the market prices and the computed ones

$$\sum_{j\in\mathcal{J}} \frac{1}{C_j^2} \left(\mathbb{E}(e^{-rT}\phi_j(\bar{S}_T)) - C_j\right)^2.$$

Let us assume that the correlation coefficients $\rho_{ij} = \rho \; \forall i \neq j$. Although this hypothesis is quite restrictive regarding the covariance structure, practitioners often assume such a structure to ensure that the problem is well defined.

As recalled above, the Γ matrix must be definite positive. First, we have to find the set of admissible values for ρ. The eigenvalues of Γ are $1-\rho$ with multiplicity 1 and $(d-1)\rho+1$ with multiplicity $d-1$. To ensure that Γ is a well defined covariance matrix, it is necessary and sufficient to choose ρ in the open interval $\mathcal{U} = \left]\frac{-1}{d-1}, 1\right[$. For simulation reasons, we prefer to consider independent Brownian motions. So, we introduce $\mathcal{W} = \{W_t, t \geq 0\}$ a $d-$dimensional standard Brownian motion. Let $\mathcal{L}$ denote the Cholesky factorisation of Γ. We know that the vector $(B^i_T)_{i=1\cdots d}$ has the same distribution as $\mathcal{L}W_T$. If we consider the process $(S_t)_{t\geq 0}$ such that $dS^i_t = S^i_t(rdt + \sigma_i\mathcal{L}_i dW_t)$ where $\mathcal{L}_i$ denotes the $i-$th line of matrix $\mathcal{L}$, then S_T and $\bar{S}_T$ are equal in distribution. Henceforth, the criteria can be rewritten using S_T

$$\sum_{j\in\mathcal{J}} \frac{1}{C_j^2} \left(\mathbb{E}(e^{-rT}\phi_j(S_T)) - C_j\right)^2. \tag{3.1}$$

We try to find the zero of the derivative of the above criteria (3.1) with respect to ρ. If we admit that we can interchange the differentiation and the expectation (see Paragraph 3.1.5 for a proof of it), we are looking for the root of the function f

$$
\begin{aligned}
f(\rho) &= \frac{1}{C^2}\,\mathbb{E}\left(e^{-rT}\frac{d\phi(S_T)}{d\rho}\right)\,(\mathbb{E}(e^{-rT}\phi(S_T)) - C),\\
f(\rho) &= \frac{1}{C^2}\,\mathbb{E}\left(e^{-rT}\nabla\phi(S_T)\cdot\nabla_\rho S_T\right)\,(\mathbb{E}(e^{-rT}\phi(S_T)) - C),\\
f(\rho) &= \frac{1}{C^2}\,\mathbb{E}\left(e^{-rT}(e^{-rT}\phi(\tilde{S}_T) - C)\,\nabla\phi(S_T)\cdot\nabla_\rho S_T\right),
\end{aligned}
$$

where $\tilde{S}_t$ is an independent copy of S_t.

The function f is defined by an expectation

$$f(\rho) = \mathbb{E}(F(\rho, G^1, G^2))$$

where G^1 and G^2 are two independent random variables with law $\mathcal{N}(0, I_{\mathbb{R}^d})$. Finding the root of such a function is precisely what stochastic algorithms are designed for.

3.1.2 Minimisation of the criteria

We will now explain how to minimise criteria (3.1) using a stochastic algorithm.

3.1.3 Case of a basket option

From now on, we consider options on a basket of assets with payoffs ϕ of the following type

$$\phi(S_T) = \left(\sum_{i=1}^{d}\lambda_i\, S_T^i \;-\; K\right)_+ .$$

Remark 3.1.1. *According to Paragraph 3.1.5, we know that for this type of payoffs, $\rho \longmapsto \phi$ and $\rho \longmapsto \frac{\partial\phi}{\partial\rho}$ are two bounded functions on any compact subset of $\mathcal{U}$.*

Formulation of the criteria

We can rewrite the payoff as a function of the correlation parameter and of a standard normal variable.

$$\psi(G,\rho) = \left(\sum_{i=1}^{d}\lambda_i x_i \exp\left((r - \sigma_i^2/2)T \;+\; \sigma_i\,\sqrt{T}\,\mathcal{L}_i\, G\right) - K\right)_+ .$$

For the rest of this chapter, we only consider one option in the criteria (3.1). The criteria on function f can then be rewritten

$$\begin{aligned}
f(\rho) &= \frac{1}{C^2}\,\mathbb{E}\left(e^{-rT}(e^{-rT}\phi(\bar{S}_T) - C)\left(\sum_{i=2}^{d} \lambda_i\,\sigma_i\,S_T^i\,d\mathcal{L}_i\,W_T \mathbf{1}_{\{\phi(S_T)>0\}}\right)\right),\\
f(\rho) &= \frac{1}{C^2}\,\mathbb{E}\Bigg((e^{-rT}\psi(G^{(1)},\rho) - C)\\
&\qquad \left(\sum_{i=2}^{d} \lambda_i\,\sigma_i\,S_0^i e^{-\sigma_i^2 T/2+\sigma_i\,\mathcal{L}_i(\rho)\,\sqrt{T}\,G^{(2)}}\,d\mathcal{L}_i(\rho)\sqrt{T}\,G^{(2)}\mathbf{1}_{\{\psi(G^{(2)},\rho)>0\}}\right)\Bigg),\\
&= \frac{1}{C^2}\mathbb{E}\left(F(\rho, G^{(1)}, G^{(2)})\right). \qquad (3.2)
\end{aligned}$$

where $G^{(1)}$ and $G^{(2)}$ are two independent standard normal variables in $\mathbb{R}^d$.

Our problem is a prime example of the usage of a constrained stochastic algorithm. To compute an approximation of $\rho^\star$ such that $f(\rho^\star) = 0$, we can use the algorithm defined by (1.5). More precisely, we consider $(G_n^{(1)})_{n\geq 0}$ and $(G_n^{(2)})_{n\geq 0}$ two sequences of independent standard normal variables in $\mathbb{R}^d$ and we introduce the sequence ρ_n defined by recursion for any arbitrary ρ_0 in $\mathcal{U}$.

$$\rho_{n+1} \;=\; \Pi_{\mathcal{U}_\varepsilon}(\rho_n - \gamma_{n+1}F(\rho_n, G_{n+1}^{(1)}, G_{n+1}^{(2)}))$$

where $\mathcal{U}_\varepsilon$ is a compact set strictly included in $\mathcal{U}$ such that $d(\mathcal{U}_\varepsilon, \partial\mathcal{U}) \geq \varepsilon$.

Importance Sampling

Thanks to Equation (3.2), f can easily be written as a sum of expectations.

$$f(\rho) = \sum_{i=2}^{d} \mathbb{E}(F_i(\rho, G^{(1)}, G^{(2)})).$$

For i in $\{2, \ldots, d\}$, we define $\bar{F}_i$

$$\bar{F}_i(\rho, G^{(1)}, G^{(2)}, \mu_i) = F_i(\rho, G^{(1)}, G^{(2)} + \mu)e^{-\mu_i\cdot G^{(2)} - \frac{1}{2}\,\|\mu_i\|^2}, \qquad (3.3)$$

for any matrix $\mu = (\mu_1, \ldots, \mu_d)$. By using a classical change of variables in the expectation, it is quite easy to prove that $\mathbb{E}(F_i(\rho, G^{(1)}, G^{(2)})) = \mathbb{E}(\bar{F}_i(\rho, G^{(1)}, G^{(2)}))$.

To improve the behaviour of the algorithm we would like to use functions $\bar{F}_i$ with the least possible variances (or square expectations since their expectations are constant).

Proposition 3.1.2. *If in Equation (3.3) we choose*

$$\mu_i = \sigma_i \mathcal{L}_i \sqrt{T},$$

then f can be written

$$f(\rho) = \mathbb{E}\left(\frac{e^{-rT}\psi(G^{(1)},\rho) - C}{C^2}\sum_{i=2}^{d}\lambda_i\,\sigma_i\,S_0^i\,d\mathcal{L}_i(\rho)\,\sqrt{T}G^{(2)}\mathbf{1}_{\{\psi(G^{(2)}+\mu_i,\rho)>0\}}\right).$$

Proof. Using this value for μ_i we can rewrite Equation (3.3).

$$\begin{aligned}\bar{F}_i(\rho, G^{(1)}, G^{(2)}, \mu_i) &= \frac{e^{-rT}\psi(G^{(1)},\rho) - C}{C^2}\,\lambda_i\sigma_i S_0^i d\mathcal{L}_i(\rho)\\ &\qquad \sqrt{T}(G^{(2)} + \mu_i)\mathbf{1}_{\{\psi(G^{(2)}+\mu_i,\rho)>0\}}.\end{aligned}$$

If we split the previous expression into two different sums, we get

$$\begin{aligned}&\mathbb{E}(\bar{F}_i(\rho, G^{(1)}, G^{(2)}, \mu_i))\\ &= \mathbb{E}\left(\frac{e^{-rT}\psi(G^{(1)},\rho) - C}{C^2}\lambda_i\,\sigma_i\,S_0^i\,d\mathcal{L}_i(\rho)\,\sqrt{T}G^{(2)}\mathbf{1}_{\{\psi(G^{(2)}+\mu_i,\rho)>0\}}\right) +\\ &\quad \mathbb{E}\left(\frac{e^{-rT}\psi(G^{(1)},\rho) - C}{C^2}\right)\lambda_i\,\sigma_i\,S_0^i\,d\mathcal{L}_i(\rho)\,\mu_i\,\sqrt{T}\,\mathbb{E}\left(\mathbf{1}_{\{\psi(G^{(2)}+\mu_i,\rho)>0\}}\right).\end{aligned}$$

By definition of matrix $\mathcal{L}$, for all i in $1,\ldots,d$, $\|\mathcal{L}_i\|^2 = 1$. So ,

$$d\mathcal{L}_i(\rho)\mu_i = \frac{\sigma_i\sqrt{T}}{2}\,d(\|\mathcal{L}_i\|^2) = 0.$$

Hence, we come up with a new expression for the function f.

$$f(\rho) = \mathbb{E}\left(\frac{e^{-rT}\psi(G^{(1)},\rho) - C}{C^2}\sum_{i=2}^{d}\lambda_i\,\sigma_i\,S_0^i\,d\mathcal{L}_i(\rho)\,\sqrt{T}G^{(2)}\mathbf{1}_{\{\psi(G^{(2)}+\mu_i,\rho)>0\}}\right). \tag{3.4}$$

■

Remark 3.1.3. *Note that the drift we use is different in each expectation appearing in f. Being able to split f into a sum of expectations is definitely essential.*

Proposition 3.1.2 gives a new expression for the function f. We can use the new expression to approximate $\rho^\star$ which suggests to define a new sequence $(\bar{\rho}_n)_{n\geq 0}$.

$$\bar{\rho}_{n+1} = \begin{cases}\bar{\rho}_n - \frac{\gamma}{n+1}\,\bar{F}(\bar{\rho}_n, G^{(1)}_{n+1}, G^{(2)}_{n+1}) & \text{if } \bar{\rho}_n - \frac{\gamma}{n+1}\,\bar{F}(\bar{\rho}_n, G^{(1)}_{n+1}, G^{(2)}_{n+1}) \in \mathcal{U}_\varepsilon,\\ \bar{\rho}_n & \text{otherwise.}\end{cases} \tag{3.5}$$

3.1.4 Numerical examples

Practically, the sequence $(\gamma_n)_n$ is often chosen of the type $\left(\frac{\gamma}{n+1}\right)_n$.

Now, we present a few examples of the method described above using algorithm (3.5). The examples will be based on a basket of three assets.

- Let us consider an option on a three asset basket of payoff $(S_T^1 + S_T^2 + S_T^3 - K)_+$, with the following characteristics

$$\begin{array}{lll} S_0^1 = 30, & S_0^2 = 10, & S_0^3 = 15, \\ \sigma_1 = 0.3, & \sigma_2 = 0.4, & \sigma_3 = 0.25, \\ r = 0.035, & T = 3, & K = 90, \end{array}$$

  and with a correlation between the different assets equal to 0.3.

  Figure 3.1 shows the estimation of the correlation parameter using the method described above with $\gamma = 5$. We obtain an approximated correlation of 0.303 whereas the real value is 0.3, which gives a rather accurate estimation. The results seem good, even though the influence of the choice of the gain parameter on the accuracy of the approximation is not to be neglected. Choosing the step of the algorithm is always a bit tricky. If it is chosen too large, the remaining noise forces to increase the number of iterations. On the contrary, a too small step quickly leads the algorithm astray and in this latter case, a true numerical convergence will never be observed. The choice of parameter γ is definitely a burning issue. One way to decrease the impact of the choice of the step is to use averaging stochastic approximation as the one we studied in Chapter 2 as the comparison between the two curves on Figure 3.1 shows it.

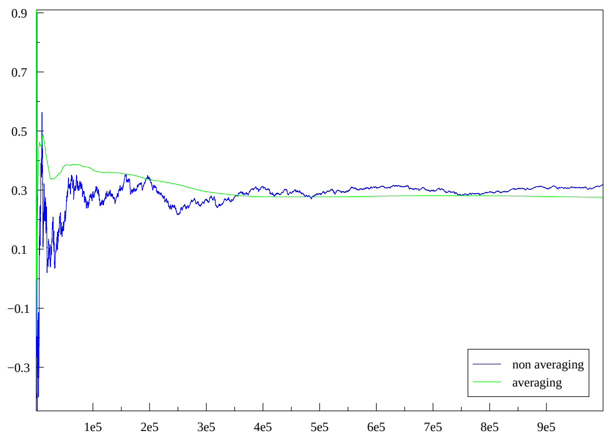

Figure 3.1: Convergence of the correlation estimate

  For any random estimator, the confidence interval is often more important than the computed value itself. The empirical cumulative distribution of the error $\rho - \rho^\star$ enables us to appreciate the quality of the convergence of the algorithm. One can also compute the probability $\mathbb{P}(\rho \in [\rho^* - \epsilon, \rho^* + \epsilon])$ (see Figure 3.2).

- Let us consider a second example still based on a basket of three assets but with a

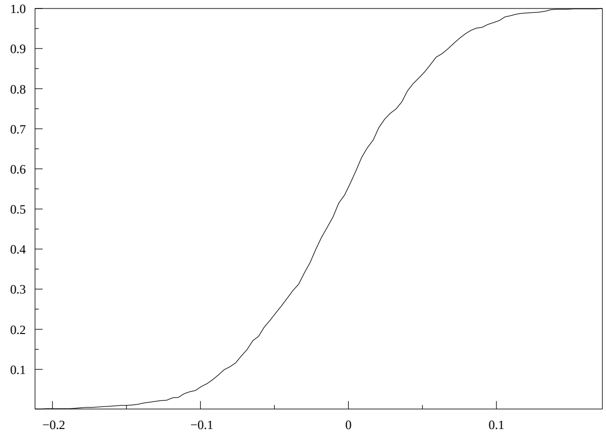

Figure 3.2: Empirical cumulative distribution function of $\rho - \rho^*$

much higher correlation equal to 0.85 this time. We consider the following payoff $(S^1_T - 5\, S^2_T - S^3_T)_+$ and the other characteristics are given below

$$\begin{array}{lll} S^1_0 = 70, & S^2_0 = 15, & S^3_0 = 35, \\ \sigma_1 = 0.3, & \sigma_2 = 0.4, & \sigma_3 = 0.45, \\ r = 0.035, & T = 2, & K = 0. \end{array}$$

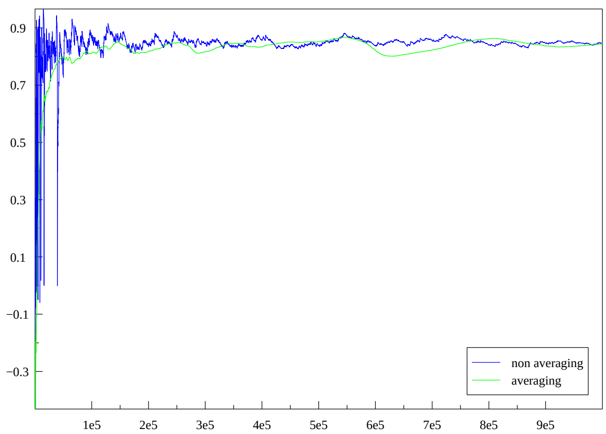

Figure 3.3: Convergence of the correlation estimate

Figure 3.4 shows that the algorithm is more precise in this case than in the previous one, anyway the averaging approximation is better once again. The confidence interval of level 90% has a length equal to 0.06, which represents an accuracy of 2%.

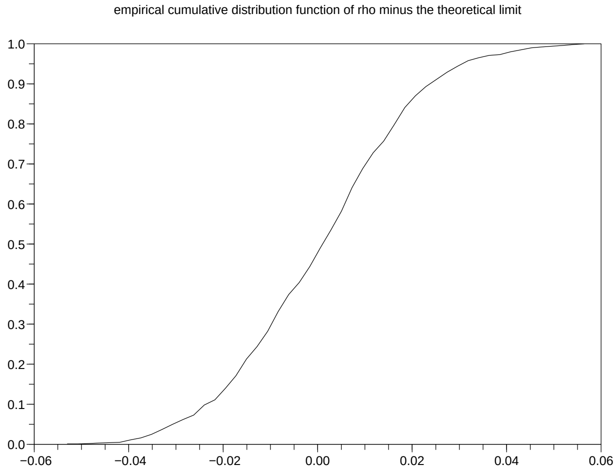

Figure 3.4: Empirical cumulative distribution function of $\rho - \rho^*$

We notice that the use of averaging procedures smooths out the numerical behaviour of the algorithm.

3.1.5 A technical condition

We show that if the payoff in (3.1) can be written as the positive part of a polynomial of d variables and of global order 1 we can differentiate within the expectation. We write $\phi(S_T) = (P(S_T))_+$ where P is a polynomial as described above. We want to prove that the following equality holds

$$\frac{\partial}{\partial \rho}\mathbb{E}(\phi(S_T)) = \mathbb{E}\left(\frac{\partial}{\partial \rho}\phi(S_T)\right).$$

$$\frac{\partial}{\partial \rho}\phi(S_T) = \begin{cases} \displaystyle\sum_{i=1}^{d} \frac{\partial}{\partial x_i}\phi(S_T)\,\frac{\partial S_T^i}{\partial \rho} & \text{if } \phi(S_T) > 0, \\ 0 & \text{otherwise.} \end{cases}$$

$\frac{\partial}{\partial x_i}\phi(S_T)$ is still a polynomial of the same order, it is sufficient to prove that $\rho \longrightarrow S_T^i(\rho)$ is bounded on $\mathcal{U}$ for all i in $1, \dots, d$ and it is the same for its derivatives.

Let us consider the Euclidean norm $\|\cdot\|$ induced on matrices. Let us compute the norm of $\mathcal{L}$.

$$\|\mathcal{L}\|^2 = \left\|\mathcal{L}\mathcal{L}^T\right\| = \|\Gamma\| = \max\left(1-\rho, (d-1)\rho + 1\right)$$

according to the computation of the eigenvalues of $\mathcal{L}$, hence $1 \leq \|\mathcal{L}\|^2 \leq d$.
Let us point out that $S_T^i = C^{ste} \exp\left(\sum_{j=1}^{i} \sigma_i \sqrt{T} \mathcal{L}_{ij} W_T^j\right)$ where C^{ste} is independent of ρ and that using the previous upper bound

$$\left(\sum_{j=1}^{i} \mathcal{L}_{ij} W_T^j\right)^2 \leq \|\mathcal{L} W_T\|^2 \leq \|\mathcal{L}\|^2 \|W_T\|^2 \leq d \|W_T\|^2 .$$

So, we can state that $\rho \longrightarrow S_T^i(\rho)$ is bounded on $\mathcal{U}$ for all i in $1, \ldots, d$. Let us handle the case of the derivative of $\mathcal{L}$.

$$\left|\frac{\partial S_T^i}{\partial \rho}\right| \leq S_T^i \, \sigma_i \, \|d\mathcal{L}\| \, \|W_T\| .$$

The eigenvalues of $d\mathcal{L}$ are the derivatives of the eigenvalues of $\mathcal{L}$ since the latter is a triangular matrix.

$$\|d\mathcal{L}\| = \max\left(\frac{1}{2\sqrt{1-\rho}}, \frac{d-1}{2\sqrt{(d-1)\rho+1}}\right).$$

Hence, we have shown that if we are on a compact subset of $\mathcal{U}$, the norm of the derivative of $\mathcal{L}$ remains bounded. Lebesgue's theorem allows us to interchange the expectation and the differentiation.

3.2 A Variance reduction technique

In this section, we use truncated stochastic approximations to implement a variance reduction for Monte Carlo simulations. This is based on the work of Arouna [6, 7]. Suppose you want to compute $\mathbb{E}(X)$ where X is a real-valued r.v. and suppose you can construct a family of r.v. $(X_\theta, \theta \in \mathbb{R})$ such that $\mathbb{E}(X_\theta) = \mathbb{E}(X)$. Then, if $\mathbb{E}(X)$ is to be computed using Monte Carlo simulations, it is quite natural to try to find the value $\theta^\star$ of the parameter θ that minimises $\mathrm{Var}(X_\theta)$. The idea of Arouna is to use this technique in the background of option pricing and to compute an estimator of $\theta^\star$ using a truncated stochastic approximation.

First, we present the problem and explain how to construct such a family $(X_\theta)_\theta$ in a Gaussian framework. Then, we propose to estimate $\theta^\star$ using an averaging truncated stochastic algorithm. We also extend the convergence result obtained by Arouna [7] for the estimator and prove that the considered estimator is asymptotically normal. Afterwards, we discuss the different ways of implementing the variance reduction procedure together with the Monte Carlo computation. To conclude the theoretical part of this application, we study the convergence rate of the joint distribution of the estimator of $\theta^\star$ and the running Monte Carlo summation. Finally, we present some numerical results in the case of the pricing of Mountain range options and basket options.

3.2.1 Presentation of the problem

Let us consider a financial market model defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with d underlying assets $(S^i)_{i=1\dots d}$. We consider a finite time horizon $[0, T]$. $\mathbb{P}$ is a martingale probability measure. $r(t)$ denotes the instantaneous risk-free interest rate, which is supposed to be deterministic. We assume that under the risk-neutral measure $\mathbb{P}$, the dynamics of the underlying asset S^i is given by

$$dS_t^i = S_t^i(r(t)dt + \sigma_i(t, S_t^i)dW_t^i), \quad S_0^i = x^i, \tag{3.6}$$

where $\sigma_i(t, y)$ is the volatility function of the asset i and W is a $d-$dimensional Brownian motion with bracket given by $d\langle W_i, W_j\rangle_t = \rho_{i,j}dt$ for $i \neq j$. We define the correlation matrix Γ by $\Gamma_{i,j} = \rho_{i,j}$ for $i \neq j$ and $\Gamma_{i,i} = 1$. x is the vector of initial values.

In this context, the price at time T of an option with payoff $\psi(S_t; t \leq T)$ is defined as

$$p = \mathbb{E}\left(\mathrm{e}^{-\int_0^T r(t)dt}\, \psi(S_t; t \leq T)\right). \tag{3.7}$$

For the sake of simplicity, we assume from now on that the discount factor, $\mathrm{e}^{-\int_0^T r(t)dt}$, has been absorbed into the payoff function ψ. Most of the time, this expectation is computed using Monte Carlo simulations on a time discretisation grid. Then, the expectation in (3.7) only involves the value of S at times $0 = t_0 < t_1 < \cdots < t_m = T$. Using a discretisation scheme, such as the Euler scheme, for Equation (3.6), we can show that there exists a function ϕ such that p can be approximated by $\hat{p}$ on the grid $(t_0, t_1, \dots, t_m)$

$$\hat{p} = \mathbb{E}\left(\phi(W_{t_i}; 1 \leq i \leq m)\right).$$

Because $\hat{p}$ is to be computed using Monte Carlo simulations, a significant improvement of the computation relies on the use of variance reduction. Amongst all the techniques available, importance sampling is usually quite easy to implement in a Gaussian background. For a detailed exposition of the different reduction techniques commonly used in financial mathematics, we refer the reader to the book of Glasserman [34].

Using Girsanov's Theorem, (see the book of Lamberton and Lapeyre [47] for an elementary version of this theorem) $\hat{p}$ can be rewritten

$$\hat{p} = \mathbb{E}\left(\phi(W_{t_i} + \int_0^{t_i} \lambda_u du; 1 \leq i \leq m)\, \mathrm{e}^{-\int_0^T \lambda_t \cdot dW_t - \frac{1}{2}\int_0^T \|\lambda_t\|^2 dt}\right), \tag{3.8}$$

where $(\lambda_t,\, 0 \leq t \leq T)$ is a measurable process such that $\int_0^T \|\lambda_t\|^2 dt < \infty$ $a.s.$ and the process $L_t = \mathrm{e}^{-\int_0^t \lambda_u \cdot dW_u - \frac{1}{2}\int_0^t \|\lambda_u\|^2 du}$ is a martingale.

The idea is to minimise the variance of $\phi(W_{t_i} + \int_0^{t_i} \lambda_u du; 0 \leq i \leq m)L_T$ over the set of measurable processes λ such that L is a martingale. Obviously, this is completely unrealistic and we have to narrow the class of changes of probability we consider. In this work, we restrict to constant λ processes and set $\lambda_t = \theta$, where $\theta \in \mathbb{R}^d$. This choice corresponds to adding a linear drift to the Brownian motion and hence changing its mean.

With this choice for the process λ, one can rewrite (3.8)

$$\hat{p} = \mathbb{E}\left(\phi(W_{t_i} + \theta t_i; 1 \le i \le m)\, \mathrm{e}^{-\theta \cdot W_T - \frac{\|\theta\|^2 T}{2}}\right). \tag{3.9}$$

Now, we focus on the minimisation of the variance of $\phi(W_{t_i} + \theta t_i; 0 \le i \le m)\, \mathrm{e}^{-\theta \cdot W_T - \frac{\|\theta\|^2 T}{2}}$. In fact, since the expectation is constant, it suffices to consider the second moment to perform the minimisation

$$v(\theta) = \mathbb{E}\left(\phi(W_{t_i} + \theta t_i; 1 \le i \le m)^2\, \mathrm{e}^{-2\theta \cdot W_T - \|\theta\|^2 T}\right). \tag{3.10}$$

Using Girsanov's Theorem again enables to remove the dependency on θ inside the function ϕ to get

$$v(\theta) = \mathbb{E}\left(\phi(W_{t_i}; 1 \le i \le m)^2\, \mathrm{e}^{-\theta \cdot W_T + \frac{\|\theta\|^2 T}{2}}\right). \tag{3.11}$$

Proposition 3.2.1. *Assume that*

- *there exists $\varepsilon > 0$ such that $\mathbb{E}(\phi(W_{t_i}; 1 \le i \le m)^{2+\varepsilon}) < \infty$,*
- *there exists a compact subset $A \subset \mathbb{R}^{d \times m}$ of strictly positive Lebesgues measure such that*

$$\exists \underline{m}, \overline{m} > 0 \ s.t. \forall x \in A, \underline{m} \le \phi(x) \le \overline{m}. \tag{3.12}$$

Then v is of class $\mathcal{C}^2$ on $\mathbb{R}^d$ and is strictly convex. Moreover,

$$\nabla v(\theta) = \mathbb{E}\left((\theta T - W_T)\phi(W_{t_i}; 1 \le i \le m)^2\, \mathrm{e}^{-\theta \cdot W_T + \frac{\|\theta\|^2 T}{2}}\right). \tag{3.13}$$

Proof. For the sake of clearness, we do the proof in the case $\theta \in \mathbb{R}$. Let us consider $f(\theta, W) = (\theta T - W_T)\phi(W_{t_i}; 0 \le i \le m)^2\, \mathrm{e}^{-\theta \cdot W_T + \frac{\|\theta\|^2 T}{2}}$. For θ in any compact set $[a, b]$ with $a < b$, we have

$$|f(\theta, W)| \le \phi(W_{t_i}; 1 \le i \le m)^2\, \mathrm{e}^{c^2/2T}\, \mathrm{e}^{c|W_T|}(|W_T| + c)$$

where $c = \max(|a|, |b|)$. Using Hölder's inequality, we have for any $p, q > 0$ s.t. $\frac{1}{p} + \frac{1}{q} = 1$

$$\begin{aligned}\mathbb{E}\left(\phi(W_{t_i}; 1 \le i \le m)^2\, \mathrm{e}^{c^2/2T}\, \mathrm{e}^{c|W_T|}(|W_T| + c)\right) \le \\ \mathrm{e}^{c^2/2T}\, \mathbb{E}\left(\phi(W_{t_i}; 1 \le i \le m)^{2p}\right)^{1/p} \mathbb{E}\left(\mathrm{e}^{qc|W_T|}(|W_T| + c)^q\right)^{1/q}.\end{aligned}$$

For any $q > 0$, the second expectation on the right hand side is finite and if we choose $p = 1 + \varepsilon/2$, then $\mathbb{E}(\phi(W_{t_i}; 1 \le i \le m)^{2p}) < \infty$. We can use Lebesgue's theorem to interchange the expectation and the differentiation.

Hence, v is of class $\mathcal{C}^1$ on $\mathbb{R}^d$. Following the same scheme, it is quite easy to prove that v is in fact of class $\mathcal{C}^2$ and that

$$\nabla\nabla v(\theta) = \mathbb{E}\left((\theta T - W_T)(\theta T - W_T)'\phi(W_{t_i}; 1 \leq i \leq m)^2 \,\mathrm{e}^{-\theta\cdot W_T + \frac{\|\theta\|^2 T}{2}}\right) \\ + \mathbb{E}\left(T\mathcal{I}_d \phi(W_{t_i}; 1 \leq i \leq m)^2 \,\mathrm{e}^{-\theta\cdot W_T + \frac{\|\theta\|^2 T}{2}}\right).$$

Both terms in the Hessian are positive and the second is bounded from below (in the sense of the ordering relation on symmetric matrices)

$$\mathbb{E}\left(\phi(W_{t_i}; 1 \leq i \leq m)^2 \,\mathrm{e}^{-\theta\cdot W_T + \frac{\|\theta\|^2 T}{2}}\right) \geq \underline{\mathrm{m}}^2 \mathbb{E}\left(\mathbf{1}_{\{(W_{t_i}; 1 \leq i \leq m) \in A\}} \,\mathrm{e}^{-\theta\cdot W_T + \frac{\|\theta\|^2 T}{2}}\right) > 0.$$

Therefore, the Hessian of v is definite positive and v is strictly convex. ■

Corollary 3.2.2. *Under the hypotheses of Proposition 3.2.1, there is a unique* $\theta^\star \in \mathbb{R}^d$ *such that* $\min_{\theta\in\mathbb{R}^d} v(\theta) = v(\theta^\star)$.

Proof. As we already know that v is strictly convex, it is sufficient to prove that the infimum, which is unique, is actually attained for a finite value of θ.

For the sake of simplicity, we assume that W is real valued and $\theta \in \mathbb{R}$ (i.e. $d = 1$). The existence of some θ s.t. $\nabla v(\theta) = 0$ is guaranteed as soon as $\lim_{|\theta|\to\infty} \nabla\nabla v(\theta) > 0$.

$$\begin{aligned} \nabla\nabla v(\theta) &\geq \mathbb{E}\left(T\phi(W_{t_i}; 1 \leq i \leq m)^2 \,\mathrm{e}^{-\theta\cdot W_T + \frac{\|\theta\|^2 T}{2}} \mathbf{1}_{\{(W_{t_i}; 1 \leq i \leq m) \in A\}}\right) \\ &\geq T\underline{\mathrm{m}}^2 \int_A \mathrm{e}^{-\theta\sqrt{T}x + \frac{\|\theta\|^2 T}{2} - \frac{x^2}{2}}\, dx. \end{aligned}$$

For any fixed x, $\mathrm{e}^{-\theta\sqrt{T}x + \frac{\|\theta\|^2 T}{2} - \frac{x^2}{2}}$ converges to infinity as $|\theta|$ goes to infinity. Moreover, as A is a compact set, there exist a, b such that $A \subset [a, b]$. Hence, for any fixed $x \in A$, the function $\theta \longmapsto \mathrm{e}^{-\theta\sqrt{T}x + \frac{\|\theta\|^2 T}{2} - \frac{x^2}{2}}$ is increasing on $(b\sqrt{T}, +\infty)$. As A has a strictly positive Lebesgues measure, it ensues from the monotone convergence theorem that $\nabla\nabla v(\theta) \xrightarrow[\theta\to+\infty]{} +\infty$. A similar reasoning enables to prove that $\nabla\nabla v(\theta) \xrightarrow[\theta\to-\infty]{} +\infty$.

Finally, $\lim_{|\theta|\to\infty} \nabla\nabla v(\theta) = \infty$, which ensures that v attains its minimum for a finite value of θ.

■

From now on, we will assume that ϕ satisfies the assumptions of Proposition 3.2.1. The minimisation problem has been turned into a zero localisation problem.

We will show that randomly truncated algorithms are particularly well suited to approximate $\theta^\star$.

3.2.2 The procedure

In this part, we explain how the stochastic algorithm presented in Section 2.1 (see Equation (1.10)) can be used to compute the root of ∇v. ∇v can be written $\nabla v(\theta) = \mathbb{E}(U(\theta, Z))$ where

$$U(\theta, Z) = (\theta T - Z(:, m))\phi(Z)^2 \, \mathrm{e}^{-\theta \cdot Z(:,m) + \frac{\|\theta\|^2 T}{2}}.$$

with Z a r.v. with values in $\mathbb{R}^{d\times m}$ following the law of $(W_{t_i}; 1 \leq i \leq m)$.

We consider an increasing sequence of compact sets $(\mathcal{K}_j)_{j\in\mathbb{N}}$ satisfying

$$\bigcup_{j=0}^{\infty} \mathcal{K}_j = \mathbb{R}^d \text{ and } \forall j, \ \mathrm{int}(\mathcal{K}_j) \subsetneq \mathcal{K}_{j+1}. \tag{3.14}$$

Based on this sequence of compact sets, we can define

$$\begin{cases} & \theta_{n+\frac{1}{2}} = \theta_n - \gamma_{n+1} U(\theta_n, Z_{n+1}), \\ \text{if } \theta_{n+\frac{1}{2}} \in \mathcal{K}_{\sigma_n} & \theta_{n+1} = \theta_{n+\frac{1}{2}} \quad \text{and} \quad \sigma_{n+1} = \sigma_n, \\ \text{if } \theta_{n+\frac{1}{2}} \notin \mathcal{K}_{\sigma_n} & \theta_{n+1} = \theta_0 \quad \text{and} \quad \sigma_{n+1} = \sigma_n + 1. \end{cases} \tag{3.15}$$

where $(Z_n)_n$ is an i.i.d. sequence of random variables in $\mathbb{R}^{d\times m}$ following the law of Z and $\gamma_n = \frac{\gamma}{(n+1)^\alpha}$, with $1/2 < \alpha < 1$. Here, we exclude the value $\alpha = 1$ because we intend to use an averaging procedure on top of this algorithm.

Proposition 3.2.3. *If there exists $\varepsilon > 0$ such that $\mathbb{E}(\phi(Z)^{4+\varepsilon}) < \infty$ then, the sequence θ_n converges a.s. to $\theta^\star$ for any increasing sequence of compact sets $(\mathcal{K}_j)_j$ satisfying (3.14).*

Proof. To prove the convergence of $(\theta_n)_n$, it is sufficient to prove that the sequence $(\theta_n)_n$ defined by (3.15) satisfies the hypotheses of Proposition 1.2.1 (page 27).

- Under the hypothesis of Proposition 3.2.3, ∇v is strictly convex by using Proposition 3.2.1. Hence, it satisfies Hypothesis ($\mathcal{A}$1.4-*i*).
- The choice of α in $(1/2, 1)$ ensures that $\sum_n \gamma_n = +\infty$ and $\sum_n \gamma_n^2 < \infty$.
- It remains to prove that Hypothesis ($\mathcal{A}$1.7) is satisfied. To do so, we will use Corollary 1.2.2: it is sufficient to show that $\theta \longmapsto \mathbb{E}(\|U(\theta, Z)\|^2)$ is bounded on any compact set of $\mathbb{R}^d$ to ensure that Hypothesis ($\mathcal{A}$1.7) is satisfied.

  Because the proof of this last point is very similar to the proof of Proposition 3.2.1, we dare omit it. ■

Remark 3.2.4. *Proposition 3.2.3 extends the result of Arouna [7, Theorem 4]. Our result holds true for any increasing sequence of compacts sets $(\mathcal{K}_j)_j$ satisfying (3.14) whereas Arouna needed a condition on the compact sets to ensure the convergence of $(\theta_n)_n$. From a practical point of view, the weakening of the hypotheses is a great improvement.*

Proposition 3.2.5. *We assume that*

- *there exists $\varepsilon > 0$ such that $\mathbb{E}(\phi(Z)^{4+\varepsilon}) < \infty$,*
- *the sequence of compact sets satisfies (3.14) and there exists $\eta > 0$, s.t. $d(\mathcal{K}_j, \theta^\star) > \eta$ for all j.*

Then,

$$\frac{\theta_n - \theta^\star}{\sqrt{\gamma_{n+1}}} \xrightarrow[n\to\infty]{law} \mathcal{N}(0, V),$$

where

$$\begin{aligned} V &= \int_0^\infty \mathrm{e}^{-Au} \Sigma \, \mathrm{e}^{-Au} \, du, \\ A &= \nabla\nabla v(\theta^\star), \\ \Sigma &= \mathbb{E}\left((\theta^\star T - Z(:,m))(\theta^\star T - Z(:,m))' \phi(Z)^4 \, \mathrm{e}^{-2\theta^\star \cdot Z(:,m) + \|\theta^\star\|^2 T} \right). \end{aligned}$$

Proposition 3.2.5 is a corollary of Theorem 2.2.4.

Since the variance reduction we settle here aims at being automatic in the sense that it does not require any fiddling depending on the function ϕ, it is quite logic to average the procedure defined by Equation (3.15).

The averaging procedure. Based on Chapter 2, we know that on top of the procedure defined by Equation (3.15) we can add an averaging algorithm. Let $t > 0$ be the averaging window length, we define

$$\hat{\theta}_n(t) = \frac{\gamma_n}{t} \sum_{i=n}^{n+\lfloor t/\gamma_n \rfloor} \theta_i. \tag{3.16}$$

The almost sure convergence of $(\hat{\theta}_n(t))_n$ can easily be deduced from Proposition 3.2.3.

Proposition 3.2.6. *If there exists $\varepsilon > 0$ such that $\mathbb{E}(\phi(Z)^{4+\varepsilon}) < \infty$ then, the sequence $\hat{\theta}_n(t)$ converges a.s. to $\theta^\star$ for any $t > 0$ and any increasing sequence of compact sets $(\mathcal{K}_j)_j$ satisfying (3.14).*

Proposition 3.2.7. *We assume the hypotheses of Proposition 3.2.5. Then, for each $t > 0$*

$$\frac{\hat{\theta}_n(t) - \theta^\star}{\sqrt{\gamma_{n+1}}} \xrightarrow[n\to\infty]{law} \mathcal{N}(0, T),$$

where $T = \frac{A^{-1}\Sigma A}{t} + \frac{C}{t^2}$ with

$$\begin{aligned} A &= \nabla\nabla v(\theta^\star), \\ \Sigma &= \mathbb{E}\left((\theta^\star T - Z(:,m))(\theta^\star T - Z(:,m))' \phi(Z)^4 \, \mathrm{e}^{-2\theta^\star \cdot Z(:,m) + \|\theta^\star\|^2 T} \right), \\ C &= A^{-2}(\mathrm{e}^{-At} - \mathcal{I})V + VA^{-2}(\mathrm{e}^{-At} - \mathcal{I}), \\ V &= \int_0^\infty \mathrm{e}^{-Au} \Sigma \, \mathrm{e}^{-Au} \, du. \end{aligned}$$

Proof. To prove this result, we use Theorem 2.2.3 (page 47). Under the hypotheses of Proposition 3.2.7,

- ∇v is strictly convex and v is of class C^2 on $\mathbb{R}^d$, hence Hypothesis ($\mathcal{A}2.1$) is satisfied.
- Taking $\rho = \varepsilon/2$ and Σ as defined in the Proposition enables to fulfil Hypothesis ($\mathcal{A}2.2$).
- Hypothesis ($\mathcal{A}2.3$) is satisfied.

Hence, the conclusion of the Proposition ensues from Theorem 2.2.3. ■

3.2.3 Implementation of the importance sampling strategy

In this part, we assume that there exists $\varepsilon > 0$ such that $\mathbb{E}(\phi(Z)^{4+\varepsilon}) < \infty$ and that the increasing sequence of compact sets $(\mathcal{K}_j)_j$ satisfies (3.14).

There are two strategies to implement the variance reduction procedure presented above. Either one uses a first set of samples to compute an approximation of $\theta^\star$ using algorithm (3.15) or (3.16) and a new set of samples to compute $\hat{p}$ (Equation (3.9)) using a Monte Carlo method with the approximation of $\theta^\star$ found before; or one uses an adaptive strategy which means theta the same samples are used to compute the approximation of $\theta^\star$ and the Monte Carlo summation.

The non adaptive algorithm

Let $\theta_0 \in \mathbb{R}^d$ and $X_0 = 0$.

Algorithm 3.2.8. *Let n be the number of samples used for the Monte Carlo computation.*

1. *Draw a first set of samples following the law of Z to compute an estimate of $\theta^\star$, either by using $(\theta_i)_{i \leq n}$ (see Equation (3.15)) or $(\hat{\theta}_i)_{i \leq n}$ (see Equation (3.16)). We denote the computed estimate by $\tilde{\theta}$.*
2. *Draw a second set of n samples following the law of Z independent of the first set to compute*
$$X_n = \frac{1}{n} \sum_{i=1}^{n} \phi(Z_i + (\tilde{\theta} t_1, \cdots, \tilde{\theta} t_m)) \, \mathrm{e}^{-\tilde{\theta} \cdot Z_i(:,m) - \frac{\|\tilde{\theta}\|^2 T}{2}}$$

By applying Proposition 3.2.3 (resp. Proposition 3.2.6), it is clear that the sequence $(\theta_i)_i$ (resp.$(\hat{\theta}_i)_i$) defined in Algorithm 3.2.8 converges a.s. to $\theta^\star$. The convergence of $(X_i)_i$ ensues from the strong law of large numbers. Obviously, both (θ_i) (resp.$(\hat{\theta}_i)_i$) and $(X_i)_i$ satisfies a Central Limit Theorem.

The adaptive algorithm

Let $\theta_0 \in \mathbb{R}^d$ and $X_0 = 0$.

Algorithm 3.2.9. *Let n be the number of samples used for the Monte Carlo computation.*
For each i in $0, \dots, n-1$, do

1. *draw a sample Z_{i+1} according to the law of Z and independent of $\{Z_j; j \leq i\}$,*

2. *compute X_{i+1} defined by*

$$X_{i+1} = \frac{i}{i+1} X_i + \frac{1}{i+1} \phi(Z_{i+1} + (\theta_i t_1, \cdots, \theta_i t_m))\, \mathrm{e}^{-\theta_i \cdot Z_{i+1}(:,m) - \frac{\|\theta_i\|^2 T}{2}},$$

3. *compute θ_{i+1} using Equation (3.15).*

The sequence $(\theta_i)_i$ defined in Algorithm (3.2.9) converges almost surely to $\theta^\star$ by applying Proposition 3.2.3. Once the convergence of $(\theta_i)_i$ is established, the convergence of the sequence $(X_i)_i$ to $\hat{p}$ follows from Arouna [6, Theorem 1]. Moreover, it ensues from Arouna [6, Theorem 2] that the sequence $(X_i)_i$ also satisfies a Central Limit Theorem

$$\sqrt{i}(X_i - \hat{p}) \xrightarrow[i\to\infty]{law} \mathcal{N}\left(0, \mathrm{Var}\left(\phi(Z + (\theta^\star t_1, \cdots, \theta^\star t_m))\, \mathrm{e}^{-\theta^\star \cdot Z(:,m) - \frac{\|\theta^\star\|^2 T}{2}}\right)\right).$$

Note that the limiting variance appearing in the convergence above is optimal in the sense that if we had done the computation directly with $\theta^\star$ instead of an approximation we would have obtained the same limiting variance. Notice that the convergence rate of the sequence $(X_i)_i$ observed in Algorithms 3.2.9 and 3.2.8 are the same.

For the non coupled algorithm, in the case when the estimator of $\theta^\star$ is computed using Equation (3.15), we can even study the convergence rate of the couple $(\theta_i, X_i)_i$ using the results of Mokkadem and Pelletier [48].

3.2.4 A joint convergence rate

In this section, we only consider the adaptive algorithm since the convergence rate of $(\theta_i, X_i)_i$ in the non-adaptive case straightly follows from the convergence rate of each component as the samples used to compute each component are independent. We are interested in the convergence in distribution of $(\frac{\theta_n - \theta^\star}{\sqrt{\gamma_{n+1}}}, \sqrt{n}(X_n - \hat{p}))_n$ where the sequences $(\theta_n)_n$ and $(X_n)_n$ are defined by Algorithm 3.2.9. Before dealing with the general case, note that if $\gamma_n = \frac{\gamma}{n}$ (i.e. $\alpha = 1$), the sequence $(\theta_n, X_n)_n$ follows a stochastic approximation in $\mathbb{R}^{d+1}$.

the case $\alpha = 1$ In this case, the two components θ_n and X_n have the same rate of convergence. Following Equation (1.11), we can rewrite

$$\theta_{n+1} = \theta_n - \gamma_{n+1} U(\theta_n, Z_{n+1}) + \gamma_{n+1} p_{n+1} \tag{3.17}$$

where p_n is the truncating term.

We can also rewrite X_n

$$X_{n+1} = X_n + \frac{1}{n+1} \left(\phi(Z_{n+1} + (\theta_n t_1, \cdots, \theta_n t_m)) \, \mathrm{e}^{-\theta_n \cdot Z_{n+1}(:,m) - \frac{\|\theta_n\|^2 T}{2}} - X_n \right) \tag{3.18}$$

Hence, $Y_n = (\theta_n, X_n)$ satisfies a stochastic approximation in $\mathbb{R}^{d+1}$. Henceforth, the convergence in distribution of $(\frac{\theta_n - \theta^\star}{\sqrt{\gamma_{n+1}}}, \sqrt{n}(X_n - \hat{p}))_n$ ensues from the Central Limit Theorem for randomly truncated stochastic algorithms (see Theorem 1.2.5 page 32).

the case $1/2 < \alpha < 1$ When the estimator of $\theta^\star$ converges slower than the running Monte Carlo summation, the study of the joint distribution of $(\frac{\theta_n - \theta^\star}{\sqrt{\gamma_{n+1}}}, \sqrt{n}(X_n - \hat{p}))_n$ can be done in the background of the two-time scale stochastic algorithms.

In particular, from the results of Mokkadem and Pelletier [48] we can derive a Central Limit Theorem for the couple $(\frac{\theta_n - \theta^\star}{\sqrt{\gamma_{n+1}}}, \sqrt{n}(X_n - \hat{p}))_n$. The funny thing is that although θ_n and X_n are coupled, the limiting covariance matrix does not show any correlation term between θ_n and X_n. We find

$$\left(\frac{\theta_n - \theta^\star}{\sqrt{\gamma_{n+1}}}, \sqrt{n}(X_n - \hat{p}) \right)_n \xrightarrow{law} \mathcal{N}\left(0, \begin{pmatrix} \Sigma_\theta & 0 \\ 0 & \Sigma_X \end{pmatrix}\right)$$

where Σ_θ (resp. Σ_X) is the limiting variance appearing in the central limit theorem for the sequence $(\theta_n)_n$ (resp. $(X_n)_n$).

3.2.5 A few simulations

Now, we will present some numerical experiments of the procedures described above applied to the pricing of Mountain range options and basket options. We have chosen these options because they involve a large number of assets which makes the variance reduction even more challenging. The interesting point in considering Mountain range options is that their payoffs are so complex that there is no obvious direction along which the variance should decrease.

Let W be a $d-$dimensional Brownian motion such that $d\langle W_i, W_j \rangle = \rho dt$ for $i \neq j$. We recall that the multidimensional Black-Scholes model is defined by

$$\text{for } i \in \{1, \ldots, d\} \quad dS_t^i = S_t^i (r dt + \sigma_i dW_t^i), \quad S_0^i = x^i.$$

Remark 3.2.10. *For the comparison between the two algorithms to be fair, the same total number of samples is used in each algorithm, no matter how they are balanced within* *the algorithm. For one given option, the comparison between the different algorithms is performed for a fixed computational cost (the computational costs only differs between the options). For instance, when we say "we use 5000 samples for our simulation" its means that the crude Monte Carlo computation is performed with a set of 5000 samples whereas, when using variance reduction, the number of samples used by the Monte Carlo computation and the stochastic approximation is balanced such that their sum equals 5000. In this way, the computational cost of the different methods remains roughly the same.*

In the following, the size of the window for the averaging algorithm is fixed to 2. Though arbitrary this choice may seem, this is a middle of the road solution between decreasing the limiting variance (see Proposition 3.2.7) and forgetting the initial condition of the algorithm (which is the main reason not to use a standard Cesaro mean but rather o moving window one).

Application to Mountain range options

The examples we present in this part are based on the pricing of the Atlas option. We implement the different strategies described above to reduce the variance in the case of the Black-Scholes model with $r = 0.02$ and $\sigma = 0.2$.

Atlas option Consider a basket of 16 stocks, at the maturity time you remove the three best and three worst performances and pay 105% of the average performance of the remaining basket. The maturities considered are typically between 5 and 10 years. Here, we consider a maturity $T = 10$.

The comparison of Figures 3.5 and 3.6 clearly shows that the averaging algorithm has a smoother behaviour. This observation is confirmed by the comparison of the limiting variance obtained in Theorems 2.2.3 and 2.2.4. It is clear that averaging enables to provide an estimator which is asymptotically better (i.e. it has a smaller asymptotic variance). From a practical point of view, this is very important to propose an estimator with low variance so that there is no need to use a large number of samples to approximate $\theta^\star$.

Looking at Figures 3.7 and 3.8, we can see the great improvement brought by the variance reduction technique. The evolution of the strategy as described in Algorithm 3.2.8 becomes flat really quickly. Actually, the variance of the Monte Carlo summation is divided by 100 when using the variance reduction.

Application to basket options

We consider a basket of 5 assets with the following parameters $K = 200$, $T = 1$, $\sigma_i = 0.2$, $r = 0.05$, $\rho = 0.8$, $S_0 = (60, 40, 60, 30, 55)$ and $\lambda = (1, 1, 1, 1, 1)$. The payoff is

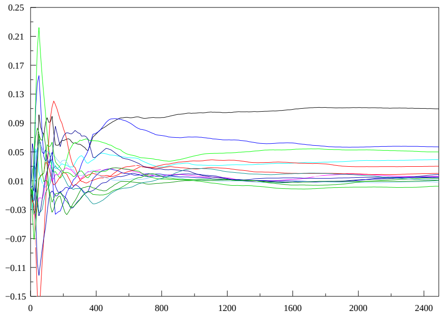

Figure 3.5: Approximation of $\theta^\star$ with averaging

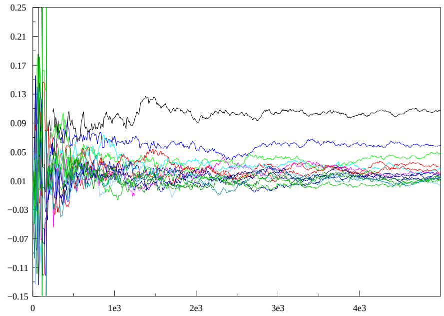

Figure 3.6: Approximation of $\theta^\star$ without averaging

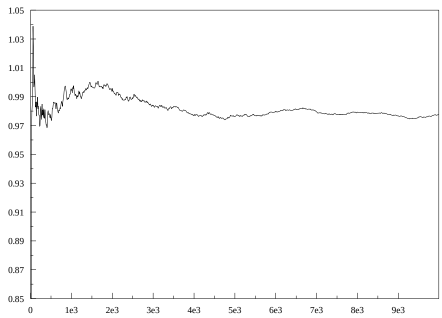

Figure 3.7: Evolution of the standard Monte Carlo simulation

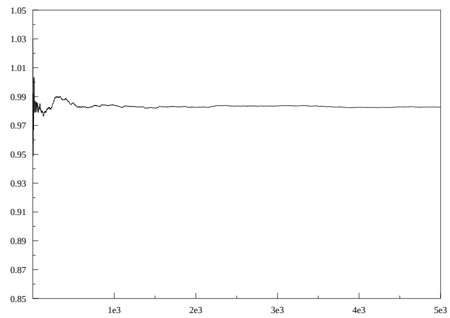

Figure 3.8: Evolution of the Monte Carlo simulation with importance sampling

given by

$$\left(\sum_{i=1}^{d} \lambda_i S_T^i - K\right)_+ .$$

We use 5000 samples for our simulations. The impression we had in the previous example is reinforced in the case of basket options. We can see that the non-averaging estimate of $\theta^\star$ (Figure 3.10) really shows a rough behaviour compared with the averaging one (Figure 3.9). Once again, the variance reduction procedure (be it Algorithm 3.2.9 or 3.2.8) has proved efficient. But, if we carefully compare Figures 3.14 and 3.12, we realise that the so promising adaptive strategy is not really competitive with the sequential algorithm that consists in first computing an estimate of $\theta^\star$ and second using this value to perform the Monte Carlo simulation.

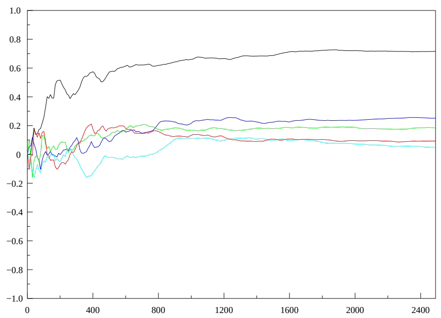

Figure 3.9: Approximation of $\theta^\star$ with averaging

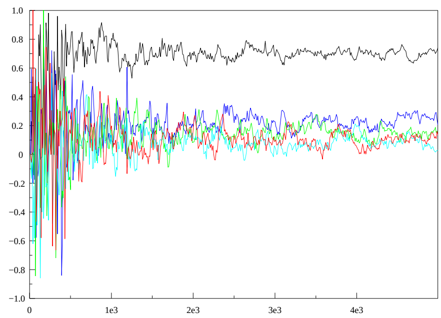

Figure 3.10: Approximation of $\theta^\star$ without averaging

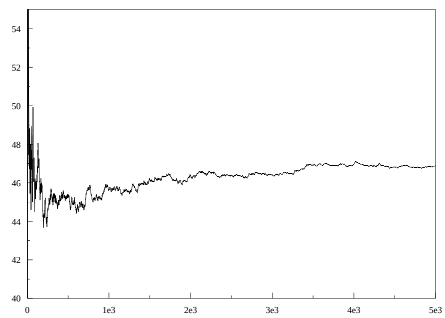

Figure 3.11: Standard Monte Carlo

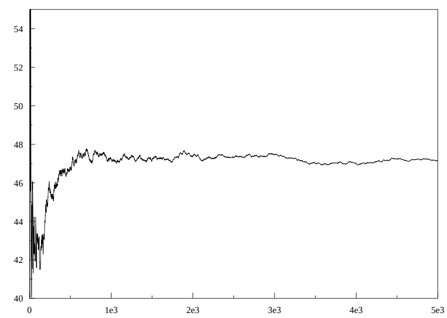

Figure 3.12: Importance Sampling coupled with MC

As we have already noticed it, the behaviour of the non averaging estimator is far less smooth than in the one of averaging algorithm. This difference highlights the

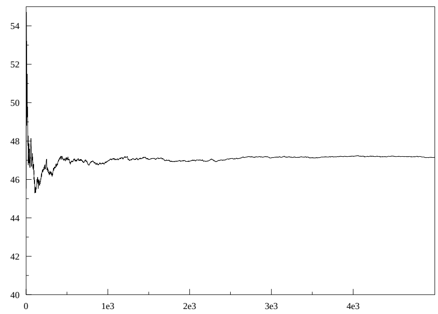

Figure 3.13: Importance Sampling + Monte Carlo (not coupled)

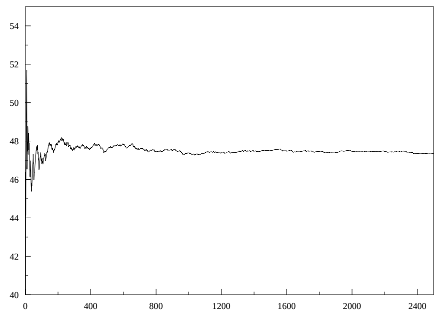

Figure 3.14: Averagging Importance Sampling with MC

high sensibility of the non averaging algorithm w.r.t the step parameter γ. This is a well-known drawback of stochastic algorithms, which is considerably reduced by using averaging as Figures 3.15 and 3.16 clearly show it. Badly choosing the step parameter is much more detrimental for the non-averaging algorithm than for the averaging one.

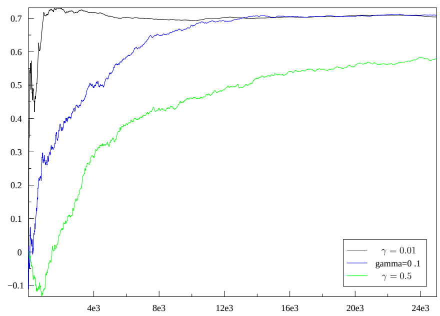

Figure 3.15: Robustness of the averaging algorithm

Finally, we show a few graphs illustrating the central limit theorem for the averaging and non-averaging estimators of $\theta^\star$. We compare the variance of the limiting variance of both estimators. Comparing Figures 3.17 and 3.18, it is crystal clear that amongst the two estimators, the averaging one has the smaller variance for the same total number of samples.

As we mentioned it earlier on, when considering the adaptive algorithm (Algorithm 3.2.9) with a step size decreasing slower than $1/n$, that is for $\alpha < 1$, the limiting correlation structure reveals two independent random variables although the estimator of $\theta^\star$ and the Monte Carlo computation are clearly coupled. The density shown in Figure 3.19 has been rebuilt using a two dimensional Epanechnikov kernel implemented in C.

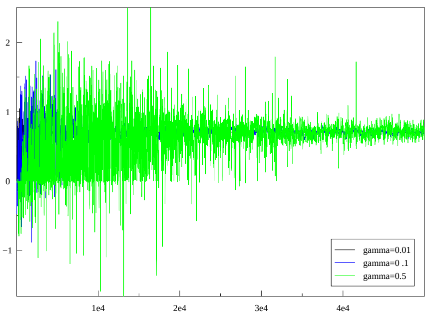

Figure 3.16: Non robustness of the non averaging algorithm

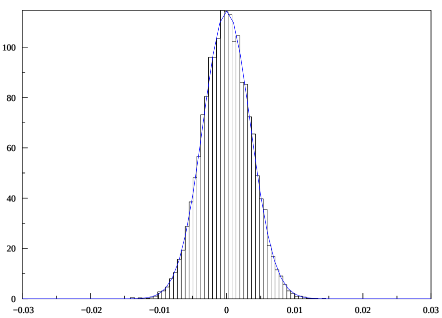

Figure 3.17: density of the averaging estimate of $\theta^\star$ for the Atlas option

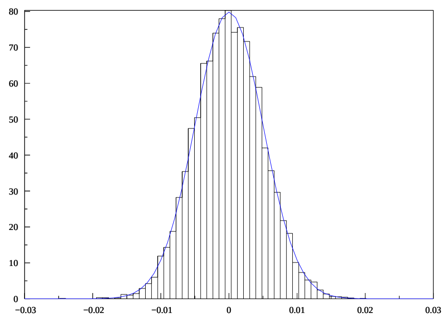

Figure 3.18: density of the non-averaging estimate of $\theta^\star$ for the Atlas option

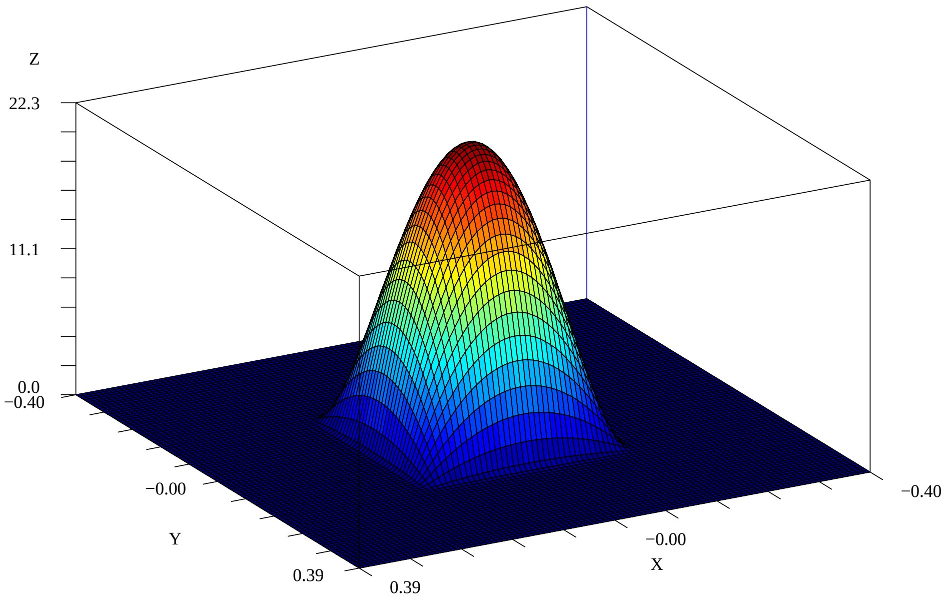

Figure 3.19: joint limiting distribution for Algorithm 3.2.9 with $\alpha = 0.9$ for the Atlas option

All the graphs presented in this part have been drawn using *Nsp* (see Chancelier et al. [20]). The code used to generate all the simulations have been written in C++. The problem we were handling was computationally highly demanding because it involved high dimensional models for instance to apply the technique to a barrier option in dimension 10 with 100 time steps we had to perform matrix-vector operations with matrices of size 10×100. Because the principle of the method is to use Monte Carlo simulations, all these complicated operations were repeated a large number of times. The matrix-vector class we designed intensively relies on the use of the *Blas* library.

Part II

Options Parisiennes

Chapter 4

Single barrier Parisian options

This chapter is based on a work article with C. Labart.

4.1 Introduction

The analysis of structured financial products often leads to the pricing of exotic options. For instance, consider a re-callable convertible bond. The holder typically wants to recall the bond if ever the underlying stock has been traded above or below a given level for too long. Such a contract can be modelled with the help of Parisian options. Parisian options are barrier options that are activated or canceled depending on the type of option if the underlying asset stays above or below the barrier long enough in a row. Parisian options are far less sensitive to influential agent on the market than standard barrier options. It is quite easy for an agent to push the price of a stock momentarily but not on a longer period so that it would affect the Parisian contract.

In this work, we study the pricing of European style Parisian options using Laplace transforms. Some other methods have already been proposed. On path dependent options, crude Monte Carlo techniques do usually not perform well. An improvement of this strategy using sharp large deviation estimates was proposed by Baldi et al. [10]. Techniques using a two dimensional partial differential equation have also drawn much attention, see for instance the works of Avellaneda and Wu [8], Haber et al. [37], or Wilmott [59]. The PDE approach is quite flexible and could even be used for American style Parisian option but the convergence is rather slow, which is badly suited for real time evaluation. In a quite similar state of mind, tree methods based on the framework of Cox et al. [26] were investigated by Costabile [25]. An original concept of implied barrier was developed by Anderluh and van der Weide [3], the idea is to replace the Parisian option by a standard barrier option with a suitably shifted barrier. The idea of using Laplace transforms to price Parisian options was introduced by Chesney et al. [23]. Their work is based on Brownian excursion theory in general and in particular on the study of the Azéma martingale (see Azéma and Yor [9]) and the Brownian meander. The prices are then computed by numerically inverting the Laplace transforms. An original way of doing so was proposed by Quittard-Pinon et al. [53]. They approximate the Laplace transforms by negative power functions whose analytical inverse is well-known. But, there is no upper bound for the error due to the inversion.

This work highly relies on the article of Chesney et al. [23] for all the theoretical results concerning the excursion theory and the way of leading the Laplace transform computations.

In this work, we give the formulae of the Laplace transforms of the prices of the different Parisian options ready to be implemented. We also derive the formulae for the prices at any time after the emission time. We prove an accuracy result for the numerical inversion of the Laplace transforms to find the prices back.

First, we define the Parisian contract and introduce some material related to the excursion theory. Then, we present a few parity relationships which enable to reduce the pricing of the eight different types of Parisian options to the pricing of the down and in call — when the barrier is smaller than the initial value — and the up and in call — when the barrier is greater than the initial value. The Laplace transforms of the prices of the two latter options are computed in Sections 4.4 and 4.5. Section 4.6 is devoted to the pricing at any time after the emission time of the option. At this stage, we are able to compute the Laplace transforms of the prices of all the different Parisian options, we only need a method to accurately invert them. In Section 4.7, we study in details the numerical inversion of Laplace transforms as introduced by Abate and Whitt [1] and prove an upper bound for the error. Finally, the last section is devoted to the comparison of our method with the enhanced Monte Carlo method of Baldi et al. [10] whose implementation in PREMIA[1] has been used for the comparison. We have also implemented our method in PREMIA.

4.2 Definitions

4.2.1 Some notations

We consider a Brownian motion $W = \{W_t, t \geq 0\}$ defined on a filtered probability space $(\Omega, \mathcal{F}, \mathbb{Q})$, which models a financial market. We assume that $\mathbb{Q}$ is the risk neutral measure and that $\mathcal{F} = (\mathcal{F}_t)_{t\geq 0}$ is the natural filtration of W. We denote by T the maturity time. In this context, we assume that the dynamics of an asset price is given by the process S

$$\forall t \in [0,T], \quad S_t = x\, \mathrm{e}^{(r-\delta-\sigma^2/2)t+\sigma W_t},$$

where $r > 0$ is the interest rate, $\delta > 0$ the dividend rate, $\sigma > 0$ the volatility and $x > 0$ the initial value of the stock. The Cameron-Martin Theorem (see Karatzas and Shreve [41]) enables to state the following proposition for a finite time horizon $[0,T]$ with $T > 0$.

[1]PREMIA is a pricing software developed by the MathFi team of INRIA Rocquencourt, see `http://www.premia.fr`.

Proposition 4.2.1. *Let* $m = \frac{1}{\sigma}\left(r - \delta - \frac{\sigma^2}{2}\right)$ *and* $\mathbb{P}$ *be a new probability, which makes* $Z = \{Z_t = W_t + mt, 0 \le t \le T\}$ *a* $\mathbb{P}$*-Brownian motion. The change of probability is given by*

$$\frac{d\mathbb{P}}{d\mathbb{Q}}_{|\mathcal{F}_T} = \mathrm{e}^{mW_T - \frac{m^2}{2}T},$$

and the dynamics of S *under* $\mathbb{P}$ *is given by*

$$\forall t \in [0, T], \quad S_t = x \ \mathrm{e}^{\sigma Z_t}.$$

Remark 4.2.2. *Since the drift term linking* W *and* Z *is deterministic,* $(\mathcal{F}_t)_{t\ge 0}$ *is also the natural filtration of* Z*.*

Before explaining what a Parisian option is, we recall the notion of excursion.

Definition 4.2.3 (Excursion)**.** For any $L > 0$ and $t > 0$, we define

$$g^S_{L,t} = \sup\{u \le t \,:\, S_u = L\} \qquad d^S_{L,t} = \inf\{u \ge t \,:\, S_u = L\}.$$

with the conventions $\sup \emptyset = 0$ and $\inf \emptyset = +\infty$. The trajectory of S between $g^S_{L,\,t}$ and $d^S_{L,\,t}$ is the excursion at level L, straddling time t.

Obviously, such an excursion can also be described in terms of the Brownian motion Z. For a given barrier L for the process S, we introduce the corresponding barrier b for Z defined by

$$b = \frac{1}{\sigma}\log\left(\frac{L}{x}\right).$$

Definition 4.2.4 (Stopping times T_b, T_b^- and T_b^+)**.** Let $b \in \mathbb{R}$ and $t > 0$, we define the hitting time of level b by

$$T_b(Z) = \inf\{u > 0 \,:\, Z_u = b\}.$$

In order to define $T_b^-(Z)$ and $T_b^+(Z)$, we introduce g^b_t and d^b_t

$$g^b_t = \sup\{u \le t \,:\, Z_u = b\}, \quad d^b_t = \inf\{u \ge t \,:\, Z_u = b\}.$$

Let $T_b^-(Z)$ denote the first time the Brownian motion Z makes an excursion longer than some fixed time D below the level b

$$T_b^-(Z) = \inf\{t > 0 \,:\, (t - g^b_t)\, \mathbf{1}_{\{Z_t < b\}} \ge D\}. \tag{4.1}$$

For the excursion above b, we define

$$T_b^+(Z) = \inf\{t > 0 \,:\, (t - g^b_t)\, \mathbf{1}_{\{Z_t > b\}} \ge D\}. \tag{4.2}$$

When no confusion is possible, we write T_b, T_b^- and T_b^+ instead of $T_b(Z)$, $T_b^-(Z)$ and $T_b^+(Z)$.

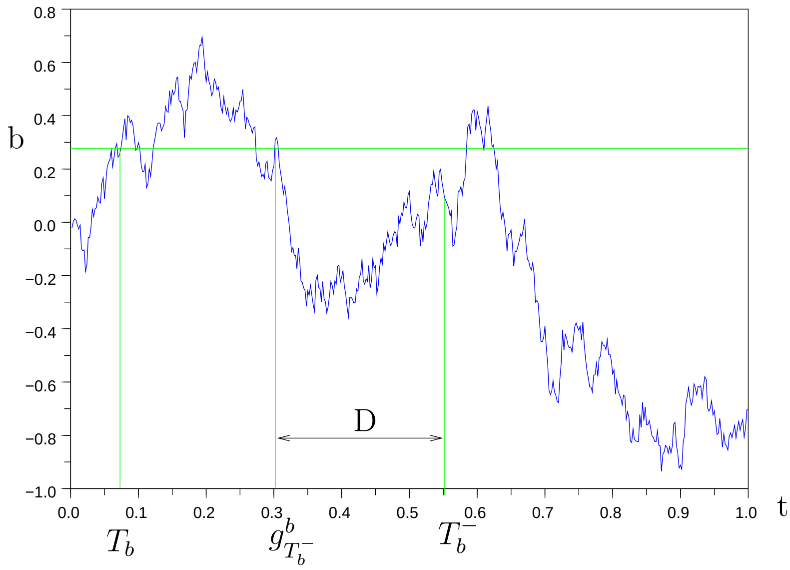

Figure 4.1: Excursion of Brownian Motion

Remark 4.2.5. *Note that* $g^b_t = g^S_{L,t}$ *and* $d^b_t = d^S_{L,t}$. *Moreover, we can also write*

$$T_b^-(Z) = \inf\{t > D \,:\, \forall s \in [t-D, t]\ Z_s \leq b\}.$$

Definition 4.2.6 (Laplace transform)**.** The Laplace Transform of a function f defined for all $t \geq 0$ is the function $\hat{f}$ defined by

$$\hat{f}(\lambda) = \int_0^{+\infty} e^{-\lambda t} f(t)dt,$$

when the integral exists.

We also recall an elementary property of the Laplace transform of the convolution of two functions

Proposition 4.2.7. *Let* f *and* g *be two functions defined on* $\mathbb{R}^+$ *whose Laplace transforms exist on* (σ_f, ∞) *and* (σ_g, ∞) *respectively, then the Laplace transform of the convolution* $f \star g$ *defined by* $(f \star g)(t) = \int_0^t g(u)f(t-u)du$ *exists on* $(\max(\sigma_f, \sigma_g), \infty)$ *and is given by*

$$\widehat{f \star g}(\lambda) = \widehat{f}(\lambda)\widehat{g}(\lambda). \tag{4.3}$$

Parisian options can be seen as barrier options where the condition involves the time spent in a row above or below a certain level and not only a hitting time. As for barrier options, which can be activated or canceled (depending whether they are In or Out) when the asset S hits the barrier, Parisian options can be activated (In options) or canceled (Out options) after S has spent more than a certain time in an excursion. Parisian options are defined in the following way

Definition 4.2.8 (Definition of θ, k and d)**.** In the following, we define

$$\theta = \sqrt{2\lambda}, \quad k = \frac{1}{\sigma}\log\left(\frac{K}{x}\right), \quad d = \frac{b-k}{\sqrt{D}}.$$

Definition 4.2.9 (Parisian Options)**.** A Parisian option is defined by three characteristics:

- Up or Down,
- In or Out,
- Call or Put.

Combining the above characteristics together enables to distinguish eight types of Parisian options. For example, PDIC denotes a Parisian Down and In call, whereas PUOP denotes a Parisian Up and Out put.

In the following section, we present Parisian Down options.

4.2.2 Parisian Down options

Parisian Down and In options

The owner of a Down and In option receives the payoff if and only if S makes an excursion below level L older than D before maturity time T. The price of a Down and In option at time 0 with payoff $\phi(S_T)$ is given by

$$\mathrm{e}^{-rT}\,\mathbb{E}_{\mathbb{Q}}\left(\phi(S_T)\mathbf{1}_{\{T_b^- < T\}}\right) = \mathrm{e}^{-(r+\frac{m^2}{2})T}\,\mathbb{E}_{\mathbb{P}}\left(\mathbf{1}_{\{T_b^- < T\}}\phi(x\,\mathrm{e}^{\sigma Z_T})\,\mathrm{e}^{mZ_T}\right). \tag{4.4}$$

For the sake of clearness, we introduce the following notation

Definition 4.2.10 (the star notation)**.** For any function f, we define

$$f^\star(t) = e^{(r+\frac{1}{2}m^2)t}f(t). \tag{4.5}$$

From (4.4), we define the price of a Parisian Down and In call.

Definition 4.2.11 (Parisian Down and In call)**.** Let $PDIC(x,T;K,L;r,\delta)$ denote the value of a Parisian Down and In call. Then,

$$PDIC(x,T;K,L;r,\delta) = \mathrm{e}^{-(r+\frac{1}{2}m^2)T}\,\mathbb{E}_{\mathbb{P}}(\mathbf{1}_{\{T_b^- < T\}}(x\,\mathrm{e}^{\sigma Z_T} - K)^+\,\mathrm{e}^{mZ_T}).$$

Using notation (4.5), we obtain

$$PDIC^\star(x,T;K,L;r,\delta) = \mathbb{E}_{\mathbb{P}}(\mathbf{1}_{\{T_b^- < T\}}(x\,\mathrm{e}^{\sigma Z_T} - K)^+\,\mathrm{e}^{mZ_T}).$$

Parisian Down and Out options

A Down and Out Parisian option becomes worthless if S reaches L and remains constantly below level L for a time interval longer than D before maturity time T. The price of a Down and Out option at time 0 with payoff $\phi(S_T)$ is given by

$$\mathrm{e}^{-rT}\,\mathbb{E}_{\mathbb{Q}}\left(\phi(S_T)\mathbf{1}_{\{T_b^->T\}}\right) = \mathrm{e}^{-(r+\frac{m^2}{2})T}\,\mathbb{E}_{\mathbb{P}}\left(\mathbf{1}_{\{T_b^->T\}}\phi(x\,\mathrm{e}^{\sigma Z_T})\,\mathrm{e}^{mZ_T}\right). \tag{4.6}$$

From (4.6), we define the price of a Parisian Down and Out call.

Definition 4.2.12 (Parisian Down and Out call)**.** Let $PDOC(x,T;K,L;r,\delta)$ denote the value of a Parisian Down and Out call. Then,

$$PDOC(x,T;K,L;r,\delta) = \mathrm{e}^{-(r+\frac{1}{2}m^2)T}\,\mathbb{E}_{\mathbb{P}}(\mathbf{1}_{\{T_b^->T\}}(x\,\mathrm{e}^{\sigma Z_T}-K)^+\,\mathrm{e}^{mZ_T}).$$

Using notation (4.5), we obtain

$$PDOC^\star(x,T;K,L;r,\delta) = \mathbb{E}_{\mathbb{P}}(\mathbf{1}_{\{T_b^->T\}}(x\,\mathrm{e}^{\sigma Z_T}-K)^+\,\mathrm{e}^{mZ_T}).$$

4.3 Relationship between prices

Parisian option prices cannot be computed directly. We are only able to give closed formulae for their Laplace transforms w.r.t. the maturity time T. As we have seen it in the above definitions, Parisian option prices depend on many parameters. The computation of the Laplace transform of one option price (say PDOC) w.r.t to T requires to distinguish several cases, depending on the relative positions of x, L and K. The sign of b $(=\frac{1}{\sigma}\log(\frac{L}{x}))$ plays an important role. In Section 4.3.2, we explain why computing the value of $\widehat{PDOC}^\star$ when $b>0$ can be reduced to computing the value of $\widehat{PDOC}^\star$ with $b=0$. As we will see it in Section 4.3.1, there also exists an In and Out parity relationship between the prices. This means that we can deduce the value of $PDOC^\star$ from the value of $PDIC^\star$. The following scheme explains how to deduce the Laplace transforms of the different Parisian call prices one from the others. Moreover, in Section 4.3.3, we state a call put parity relationship, which enables to deduce the Parisian put prices from the corresponding call prices through the Black Scholes formula.

4.3.1 In and Out parity

This part is devoted to make precise the way we compute the value of $\widehat{PDOC}^\star$ from the value of $\widehat{PDIC}^\star$. The technique developed below remains valid for Parisian Up calls. We recall Definitions 4.2.11 and 4.2.12,

$$PDIC^\star(x,T;K,L;r,\delta) = \mathbb{E}_{\mathbb{P}}(\mathbf{1}_{\{T_b^-<T\}}(x\,\mathrm{e}^{\sigma Z_T}-K)^+\,\mathrm{e}^{mZ_T}).$$
$$PDOC^\star(x,T;K,L;r,\delta) = \mathbb{E}_{\mathbb{P}}(\mathbf{1}_{\{T_b^->T\}}(x\,\mathrm{e}^{\sigma Z_T}-K)^+\,\mathrm{e}^{mZ_T}).$$

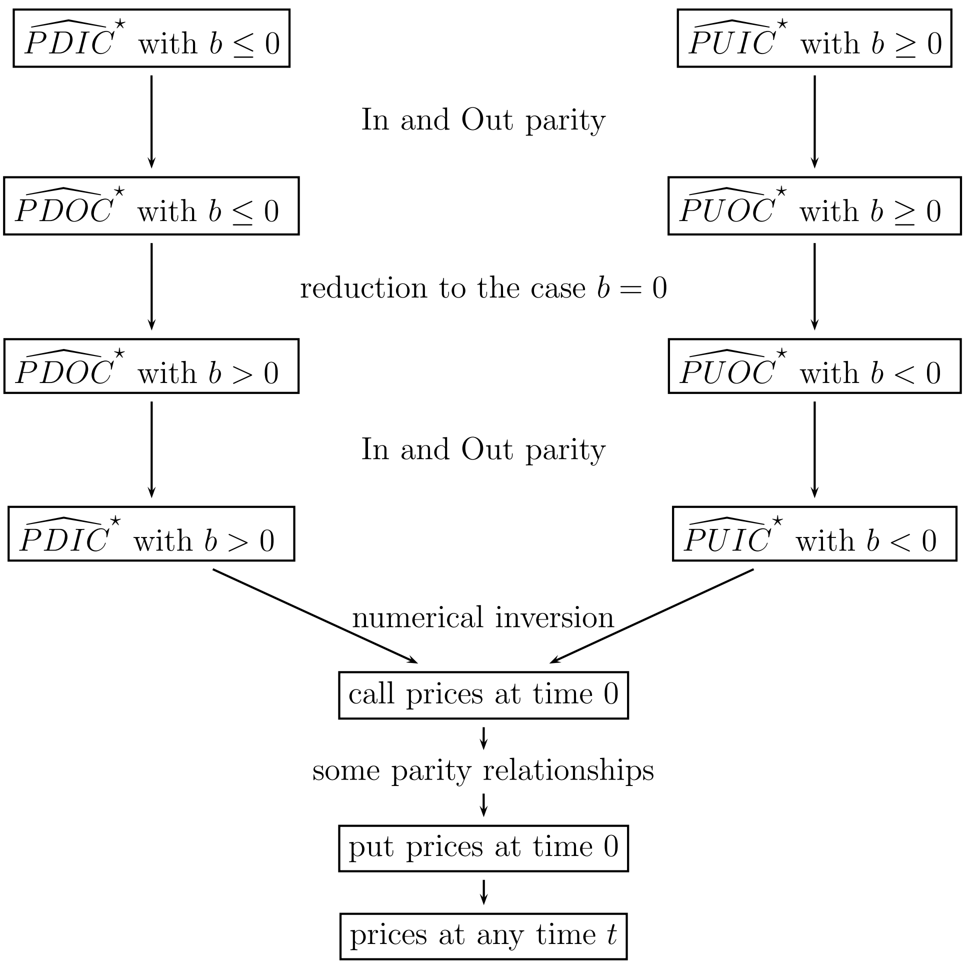

Figure 4.2: Computation scheme of Parisian option prices

By summing the two previous equalities, we get

$$PDIC^{\star}(x, T; K, L; r, \delta) + PDOC^{\star}(x, T; K, L; r, \delta) = \mathbb{E}_{\mathbb{P}}((x\,\mathrm{e}^{\sigma Z_T} - K)^{+}\,\mathrm{e}^{m Z_T}). \quad (4.7)$$

Definition 4.3.1. Let us define

$$BSC^{\star}(x, T; K; r, \delta) = \mathbb{E}_{\mathbb{P}}((x\,\mathrm{e}^{\sigma Z_T} - K)^{+}\,\mathrm{e}^{m Z_T}).$$

BSC is the price of a call option.

From (4.7), we get

$$\widehat{PDOC}^{\star}(x, \lambda; K, L; r, \delta) = \widehat{BSC}^{\star}(x, \lambda; K; r, \delta) - \widehat{PDIC}^{\star}(x, \lambda; K, L; r, \delta).$$

Then, if we manage to get closed formulae for both $\widehat{PDIC}^{\star}$ and $\widehat{BSC}^{\star}$, we can easily deduce a closed formula for $\widehat{PDOC}^{\star}$. Since the pricing of a Parisian option can only be achieved through the numerical inversion of its Laplace transform, it makes sense to compute the Laplace transform of BSC — even though it can also be accessed through the Black Scholes formula (see Black and Scholes [15]) — to be able to implement the different parity relationships straightaway.

The following proposition gives the value of $\widehat{BSC}^{\star}(x, \lambda; K; r, \delta)$

Proposition 4.3.2. *For $K \geq x$,*

$$\widehat{BSC}^{\star}(x, \lambda; K; r, \delta) = \frac{K}{\theta} \, \mathrm{e}^{(m-\theta)k} \left(\frac{1}{m-\theta} - \frac{1}{m+\sigma-\theta} \right).$$

For $K \leq x$,

$$\widehat{BSC}^{\star}(x, \lambda; K; r, \delta) = \frac{2K}{m^2 - \theta^2} - \frac{2x}{(m+\sigma)^2 - \theta^2} + \frac{K \, \mathrm{e}^{(m+\theta)k}}{\theta} \left(\frac{1}{m+\theta} - \frac{1}{m+\sigma+\theta} \right). \tag{4.8}$$

k is defined in Definition 4.2.8.

Proof. From Definition 4.3.1

$$BSC^{\star}(x, T; K; r, \delta) = \int_{-\infty}^{+\infty} \mathrm{e}^{mz} (x \, \mathrm{e}^{\sigma z} - K)^{+} \frac{1}{\sqrt{2\pi T}} \, \mathrm{e}^{-\frac{z^2}{2T}} \, dz.$$

Then,

$$\widehat{BSC}^{\star}(x, \lambda; K; r, \delta) = \int_{-\infty}^{+\infty} \mathrm{e}^{mz} (x \, \mathrm{e}^{\sigma z} - K)^{+} \int_{0}^{+\infty} \frac{\mathrm{e}^{-\lambda t}}{\sqrt{2\pi t}} \, \mathrm{e}^{-\frac{z^2}{2t}} \, dt \, dz. \tag{4.9}$$

The computation of the second integral on the right hand side is given in Appendix B.2. Combining (B.2) and (4.9), we find

$$\widehat{BSC}^{\star}(x, \lambda; K; r, \delta) = \int_{k}^{+\infty} \mathrm{e}^{mz} (x \, \mathrm{e}^{\sigma z} - K) \frac{\mathrm{e}^{-|z|\theta}}{\theta} dz. \tag{4.10}$$

- In the case $K \geq x$, $k \geq 0$ and the result easily follows.
- In the case $K \leq x$, we split the integral in (4.10) into two parts

$$\widehat{BSC}^{\star}(x, \lambda; K; r, \delta) = \int_{k}^{0} \mathrm{e}^{mz} (x \, \mathrm{e}^{\sigma z} - K) \frac{\mathrm{e}^{z\theta}}{\theta} dz + \int_{0}^{\infty} \mathrm{e}^{mz} (x \, \mathrm{e}^{\sigma z} - K) \frac{\mathrm{e}^{-z\theta}}{\theta} dz,$$

and an easy computation yields the result.

■

4.3.2 Reduction to the case $b = 0$

Assume that we know the value of $\widehat{PDOC}^{\star}$ with $b = 0$. This section aims at proving that computing $\widehat{PDOC}^{\star}$ with $b > 0$ boils down to computing the value of $\widehat{PDOC}^{\star}$ with $b = 0$, as suggested in Figure 4.2. First, we state a Proposition which links $PDOC^{\star}$ with $b > 0$ to $PDOC^{\star}$ with $b = 0$.

Proposition 4.3.3. *The price of a Parisian Down and Out call in the case $b > 0$ is given by*

$$PDOC^{\star}(x, T; K, L; r, \delta) = L\,\mathrm{e}^{mb} \int_0^D PDOC^{\star,0}(T - u; K/L; r, \delta)\mu_b(du) \tag{4.11}$$

where $\mu_b(du)$ is the law of T_b and

$$PDOC^{\star,0}(T; K; r, \delta) = \mathbb{E}_{\mathbb{P}}\left(\mathbf{1}_{\{T_0^- \geq T\}}(\mathrm{e}^{\sigma Z_T} - K)^+\,\mathrm{e}^{mZ_T}\right).$$

Remark 4.3.4. *Note that $PDOC^{\star,0}(T; K; r, \delta) = PDOC(1, T; K, 1; r, \delta)$.*

Proof. First, we recall the value of $PDOC^{\star}$

$$PDOC^{\star}(x, T; K, L; r, \delta) = \mathbb{E}_{\mathbb{P}}(\mathbf{1}_{\{T_b^- \geq T\}}(x\,\mathrm{e}^{\sigma Z_T} - K)^+\,\mathrm{e}^{mZ_T}).$$

Since Z starts from 0 and b is positive, $T_b < D$ on the set $\{T_b^- \geq T\}$. In fact, if T_b were strictly greater than D, it would mean that Z would not have crossed b before D and then T_b^- would be equal to D, which is impossible since we are on the set $\{T_b^- \geq T\}$, and $T > D$. Therefore, we can write

$$PDOC^{\star}(x, T; K, L; r, \delta) = \mathbb{E}_{\mathbb{P}}(\mathbf{1}_{\{T_b^- \geq T\}}\mathbf{1}_{\{T_b \leq D\}}(x\,\mathrm{e}^{\sigma Z_T} - K)^+\,\mathrm{e}^{mZ_T}).$$

Introducing Z_{T_b}, we can also write

$$PDOC^{\star}(x, T; K, L; r, \delta) = \\ \mathbb{E}_{\mathbb{P}}\left(\mathbf{1}_{\{T_b \leq D\}}\mathbb{E}_{\mathbb{P}}[\mathbf{1}_{\{T_b^- - T_b \geq T - T_b\}}(x\,\mathrm{e}^{\sigma Z_T - Z_{T_b} + b} - K)^+\,\mathrm{e}^{m(Z_T - Z_{T_b} + b)} \mid \mathcal{F}_{T_b}]\right).$$

To compute the inner expectation in the previous formula, we rely on the strong Markov property. Let $B = \{B_t = Z_{T_b + t} - Z_{T_b}, t \geq 0\}$. B is independent of $\mathcal{F}_{T_b}$ and one can easily prove that $T_b^-(Z) - T_b(Z) = T_0^-(B)$ a.s. on the set $\{T_b^- \geq T\}$.

Hence, we find

$$PDOC^{\star}(x, T; K, L; r, \delta) = \mathbb{E}[\mathbf{1}_{\{T_b \leq D\}}\mathbb{E}[\mathbf{1}_{\{T_0^- \geq T - t\}}(x\,\mathrm{e}^{\sigma(B_{T-t} + b)} - K)^+\,\mathrm{e}^{m(B_{T-t} + b)}]_{|t = T_b}].$$

We get

$$PDOC^{\star}(x, T; K, L; r, \delta) = \int_0^D \mathbb{E}_{\mathbb{P}}\left(\mathbf{1}_{\{T_0^- \geq T - u\}}(x\,\mathrm{e}^{\sigma(B_{T-u} + b)} - K)^+\,\mathrm{e}^{m(B_{T-u} + b)}\right)\mu_b(du),$$

where $\mu_b(du)$ is the law of T_b. As $b = \frac{1}{\sigma}\ln\left(\frac{L}{x}\right)$, we get

$$PDOC^{\star}(x,T;K,L;r,\delta) = L\,\mathrm{e}^{mb}\int_0^D \mathbb{E}_{\mathbb{P}}\left(\mathbf{1}_{\{T_0^- \geq T-u\}}(\mathrm{e}^{\sigma B_{T-u}} - K/L)^+\,\mathrm{e}^{mB_{T-u}}\right)\mu_b(du),$$

and the result follows. ■

By Using Proposition 4.3.3, we can state the following formula for the Laplace transform of $PDOC^{\star}(x,T;K,L;r,\delta)$.

Proposition 4.3.5. *The Laplace transform of $PDOC^{\star}$ when $b > 0$ is given by*

$$\widehat{PDOC}^{\star}(x,\lambda;K,L;r,\delta) = L\,\mathrm{e}^{mb}\,\widehat{PDOC}^{\star,0}(\lambda;K/L;r,\delta)\int_0^D \mathrm{e}^{-\lambda u}\,\mu_b(du),$$

where

$$\int_0^D \mathrm{e}^{-\lambda u}\,\mu_b(du) = \mathrm{e}^{-\theta b}\,\mathcal{N}\left(\theta\sqrt{D} - \frac{b}{\sqrt{D}}\right) + \mathrm{e}^{\theta b}\,\mathcal{N}\left(-\theta\sqrt{D} - \frac{b}{\sqrt{D}}\right).$$

Proof. From Proposition 4.3.3, we have

$$PDOC^{\star}(x,T;K,L;r,\delta) \;=\; \mathrm{e}^{-\lambda T}\;L\,\mathrm{e}^{mb}\int_0^D PDOC^{\star,0}(T-u;K/L;r,\delta)\mu_b(du)\mathbf{1}_{\{T>D\}}.$$

Using Proposition 4.2.7, it is quite easy to show that

$$\widehat{PDOC}^{\star}(x,\lambda;K,L;r,\delta) \;=\; L\,\mathrm{e}^{mb}\int_0^D \mu_b(du)\,\mathrm{e}^{-\lambda u}\;\widehat{PDOC}^{\star,0}(\lambda;K/L;r,\delta).$$

We refer the reader to Appendix B.1 for the computation of $\int_0^D \mu_b(du)\,\mathrm{e}^{-\lambda u}$. ■

4.3.3 call put parity

In this part, we explain how to deduce the put prices from the call prices using a parity relationship.

Proposition 4.3.6. *The following relationships hold*

$$PDOP(x,T;K,L;r,\delta) = xK\;PUOC\left(\frac{1}{x},T;\frac{1}{K},\frac{1}{L};\delta,r\right),$$
$$PUOP(x,T;K,L;r,\delta) = xK\;PDOC\left(\frac{1}{x},T;\frac{1}{K},\frac{1}{L};\delta,r\right),$$
$$PUIP(x,T;K,L;r,\delta) = xK\;PDIC\left(\frac{1}{x},T;\frac{1}{K},\frac{1}{L};\delta,r\right),$$
$$PDIP(x,T;K,L;r,\delta) = xK\;PUIC\left(\frac{1}{x},T;\frac{1}{K},\frac{1}{L};\delta,r\right).$$

Proof. Let us consider a Parisian Down and Out put

$$PDOP(x,T;K,L;r,\delta) = \mathbb{E}\left(\mathrm{e}^{mZ_T}(K - x\,\mathrm{e}^{\sigma Z_T})^+ \mathbf{1}_{\{T_b^- > T\}}\right) \mathrm{e}^{-\left(r+\frac{m^2}{2}\right)T}.$$

One notices that the first time the Brownian motion Z makes an excursion below b longer than D is equal to the first time the Brownian motion $-Z$ makes above $-b$ an excursion longer than D. Therefore, by introducing the new Brownian motion $W = -Z$, we can rewrite

$$\begin{aligned} PDOP(x,T;K,L;r,\delta) &= \mathbb{E}\left(\mathrm{e}^{-mW_T}(K - x\,\mathrm{e}^{-\sigma W_T})^+ \mathbf{1}_{\{T_{-b}^+ > T\}}\right) \mathrm{e}^{-\left(r+\frac{m^2}{2}\right)T}, \\ &= xK\,\mathbb{E}\left(\mathrm{e}^{-(m+\sigma)W_T}\left(\frac{1}{x}\,\mathrm{e}^{\sigma W_T} - \frac{1}{K}\right)^+ \mathbf{1}_{\{T_{-b}^+ > T\}}\right) \mathrm{e}^{-\left(r+\frac{m^2}{2}\right)T}. \end{aligned}$$

Let us introduce $m' = -(m+\sigma)$, $\delta' = r$, $r' = \delta$ and $b' = -b$. With these relations, we can easily check that $m' = \frac{1}{\sigma}\left(r' - \delta' - \frac{\sigma^2}{2}\right)$ and that $r' + \frac{m'^2}{2} = r + \frac{m^2}{2}$. Moreover, we notice that the barrier L' corresponding to $b' = -b$ is $\frac{1}{L}$.

Therefore, $\mathbb{E}\left(\mathrm{e}^{-(m+\sigma)W_T}\left(\frac{\mathrm{e}^{\sigma W_T}}{x} - \frac{1}{K}\right)^+ \mathbf{1}_{\{T_{-b}^+ > T\}}\right) \mathrm{e}^{-\left(r+\frac{m^2}{2}\right)T}$ is in fact the price of an Up and Out call $PUOC\left(\frac{1}{x}, T; \frac{1}{K}, \frac{1}{L}; \delta, r\right)$. Finally, we come up with the following relation

$$PDOP(x,T;K,L;r,\delta) = xK\;PUOC\left(\frac{1}{x}, T; \frac{1}{K}, \frac{1}{L}; \delta, r\right).$$

The three other assertions in Proposition 4.3.6 can be proved in the same way. ■

4.4 Valuation of Parisian calls

Looking at Figure 4.2, we notice that we only need to compute $\widehat{PDIC}^{\star}$ with $b \leq 0$ and $\widehat{PUIC}^{\star}$ with $b \geq 0$. With these values we can deduce the prices of all the other Parisian calls.

4.4.1 The valuation of a Parisian Down and In call with $b \leq 0$

Before computing the Laplace transform of $PDIC^{\star}$, we state some preliminary results. We give a new expression for $PDIC^{\star}$ in Proposition 4.4.1 and we state in Lemma 4.4.4 a key result for the computation of $\widehat{PDIC}^{\star}$.

Preliminary results

Proposition 4.4.1.

$$PDIC^{\star}(x,T;K,L;r,\delta) = \int_k^{\infty} e^{my}(xe^{\sigma y} - K)h_b^-(T,y)dy,$$

where

$$h_b^-(t,y) = \int_{-\infty}^{b} \frac{b-z}{D}\, \mathrm{e}^{-\frac{(z-b)^2}{2D}}\, \gamma^-(t,z-y)dz,$$

and

$$\gamma^-(t,x) = \mathbb{E}_{\mathbb{P}}\left(\mathbf{1}_{\{T_b^- < t\}} \frac{\mathrm{e}^{-\frac{x^2}{2(t-T_b^-)}}}{\sqrt{2\pi(t-T_b^-)}}\right). \tag{4.12}$$

Proof. Remember that the value of $PDIC^{\star}$ is given by

$$PDIC^{\star}(x,T;K,L;r,\delta) \;=\; \mathbb{E}_{\mathbb{P}}(\mathbf{1}_{\{T_b^- < T\}}(x\,\mathrm{e}^{\sigma Z_T} - K)^+\, \mathrm{e}^{m Z_T}).$$

By conditioning with respect to $\mathcal{F}_{T_b^-}$, we can write

$$PDIC^{\star}(x,T;K,L;r,\delta) = \mathbb{E}_{\mathbb{P}}(\mathbf{1}_{\{T_b^- < T\}}\mathbb{E}_{\mathbb{P}}[x\,\mathrm{e}^{\sigma(Z_T - Z_{T_b^-} + Z_{T_b^-})} - K)^+\, \mathrm{e}^{m(Z_T - Z_{T_b^-} + Z_{T_b^-})}\,|\mathcal{F}_{T_b^-}]).$$

First, we deal with the conditional expectation. Let $B_t = Z_{t+T_b^-} - Z_{T_b^-}$ for $t \geq 0$. B is independent of $\mathcal{F}_{T_b^-}$. So, we obtain

$$\mathbb{E}_{\mathbb{P}}\left[(x\,\mathrm{e}^{\sigma(Z_T - Z_{T_b^-} + Z_{T_b^-})} - K)^+\, \mathrm{e}^{m(Z_T - Z_{T_b^-} + Z_{T_b^-})}\,|\mathcal{F}_{T_b^-}\right] = \mathbb{E}_{\mathbb{P}}\left[(x\,\mathrm{e}^{\sigma(B_{T-\tau}+z)} - K)^+\, \mathrm{e}^{m(B_{T-\tau}+z)}\right]_{|z = Z_{T_b^-},\, \tau = T_b^-},$$

and

$$\mathbb{E}_{\mathbb{P}}\left[(x\,\mathrm{e}^{\sigma(B_{T-\tau}+z)} - K)^+\, \mathrm{e}^{m(B_{T-\tau}+z)}\right] = \frac{1}{\sqrt{2\pi(T-\tau)}}\left(\int_{-\infty}^{\infty} \mathrm{e}^{mu}(x\,\mathrm{e}^{\sigma u} - K)^+\, \mathrm{e}^{-\frac{(u-z)^2}{2(T-\tau)}}\, du\right).$$

Hence, we get

$$PDIC^{\star}(x,T;K,L;r,\delta) = \mathbb{E}_{\mathbb{P}}(\mathbf{1}_{\{T_b^- < T\}}\mathcal{P}_{T-T_b^-}(f_x)(Z_{T_b^-})),$$

with

$$f_x(z) = \mathrm{e}^{mz}(x\,\mathrm{e}^{\sigma z} - K)^+, \tag{4.13}$$

and

$$\mathcal{P}_t(f_x)(z) = \frac{1}{\sqrt{2\pi t}} \int_{-\infty}^{\infty} f_x(u) \exp\left(-\frac{(u-z)^2}{2t}\right) du. \tag{4.14}$$

As established by Chesney et al. [23], the random variables $Z_{T_b^-}$ and T_b^- are independent. Denoting the law of $Z_{T_b^-}$ by $\nu^-(dz)$ leads to

$$\begin{aligned} PDIC^\star(x,T;K,L;r,\delta) &= \int_{-\infty}^{\infty} \mathbb{E}_{\mathbb{P}}(\mathbf{1}_{\{T_b^-<T\}}\mathcal{P}_{T-T_b^-}(f_x)(z))\nu^-(dz), \\ &= \int_{-\infty}^{\infty} f_x(y)h_b^-(T,y)dy, \end{aligned}$$

where

$$h_b^-(t,y) = \int_{-\infty}^{\infty} \mathbb{E}_{\mathbb{P}}\left(\mathbf{1}_{\{T_b^-<t\}}\frac{\exp\left(-\frac{(z-y)^2}{2(t-T_b^-)}\right)}{\sqrt{2\pi(t-T_b^-)}}\right)\nu^-(dz).$$

Using the expression of $\nu^-(dz)$ given in Appendix B.4, we know that

$$h_b^-(t,y) = \int_{-\infty}^{b} \frac{b-z}{D}\,\mathrm{e}^{-\frac{(z-b)^2}{2D}}\,\mathbb{E}_{\mathbb{P}}\left(\mathbf{1}_{\{T_b^-<t\}}\frac{\mathrm{e}^{-\frac{(z-y)^2}{2(t-T_b^-)}}}{\sqrt{2\pi(t-T_b^-)}}\right)dz,$$

and the result follows. ■

Definition 4.4.2. Let $\psi : \mathbb{R} \to \mathbb{R}$ denote

$$\psi(z) \triangleq \int_0^{+\infty} xe^{-\frac{x^2}{2}+zx}dx = 1 + z\sqrt{2\pi}e^{\frac{z^2}{2}}\mathcal{N}(z).$$

Remark 4.4.3. *For the numerical inversion of Laplace transforms, it is important to notice that ψ is analytic on the complex plane.*

We can easily prove the following Lemma.

Lemma 4.4.4. *Let $K_{\lambda,D} : \mathbb{R} \to \mathbb{R}$ be defined as*

$$K_{\lambda,D}(a) = \int_0^{+\infty} v\,\mathrm{e}^{-\frac{v^2}{2D}-|a-v|\theta}\,dv. \tag{4.15}$$

Then,

$$K_{\lambda,D}(a) = \begin{cases} e^{\theta a}D\psi(-\theta\sqrt{D}) & \textit{if } a \le 0, \\ e^{-\theta a}D\psi(\theta\sqrt{D}) - D\theta\sqrt{2\pi D}e^{\lambda D}\Big\{\mathcal{N}(\theta\sqrt{D}-\frac{a}{\sqrt{D}})e^{-\theta a}+ & \\ \qquad\qquad \mathcal{N}(-\theta\sqrt{D}-\frac{a}{\sqrt{D}})e^{\theta a}\Big\} & \textit{otherwise.} \end{cases}$$

The Laplace transform of $PDIC^{\star}(x,T;K,L;r,\delta)$

Theorem 4.4.5. *The value of* $\widehat{PDIC}^{\star}$ *is given by the following formula*

$$\widehat{PDIC}^{\star}(x,\lambda;K,L;r,\delta)=\frac{\mathrm{e}^{\theta b}}{D\theta\psi(\theta\sqrt{D})}\int_k^{\infty}e^{my}(xe^{\sigma y}-K)K_{\lambda,D}(b-y)dy. \qquad (4.16)$$

For any $\lambda>\frac{(m+\sigma)^2}{2}$ *and for* $K>L$*, we get*

$$\widehat{PDIC}^{\star}(x,\lambda;K,L;r,\delta)=\frac{\psi(-\theta\sqrt{D})\,\mathrm{e}^{2b\theta}}{\theta\psi(\theta\sqrt{D})}K\,\mathrm{e}^{(m-\theta)k}\left(\frac{1}{m-\theta}-\frac{1}{m+\sigma-\theta}\right), \qquad (4.17)$$

and for $K\le L$

$$\begin{aligned}\widehat{PDIC}^{\star}(x,\lambda;K,L;r,\delta)=&\frac{\mathrm{e}^{(m+\theta)b}}{\psi(\theta\sqrt{D})}\left(\frac{2K}{m^2-\theta^2}\left[\psi(m\sqrt{D})-m\sqrt{2\pi D}\,\mathrm{e}^{\frac{Dm^2}{2}}\mathcal{N}(m\sqrt{D}+d)\right]\right.\\ &\left.-\frac{2L}{(m+\sigma)^2-\theta^2}\left[\psi((m+\sigma)\sqrt{D})-(m+\sigma)\sqrt{2\pi D}\,\mathrm{e}^{\frac{D}{2}(m+\sigma)^2}\mathcal{N}\left((m+\sigma)\sqrt{D}+d\right)\right]\right)\\ &+\frac{K\,\mathrm{e}^{(m+\theta)k}}{\theta\psi(\theta\sqrt{D})}\left(\frac{1}{m+\theta}-\frac{1}{m+\sigma+\theta}\right)\left[\psi(\theta\sqrt{D})-\theta\sqrt{2\pi D}\,\mathrm{e}^{\lambda D}\mathcal{N}(\theta\sqrt{D}-d)\right]\\ &+\frac{\mathrm{e}^{\lambda D}\sqrt{2\pi D}}{\psi(\theta\sqrt{D})}K\,\mathrm{e}^{2b\theta}\,\mathrm{e}^{(m-\theta)k}\mathcal{N}(-d-\theta\sqrt{D})\left(\frac{1}{m+\sigma-\theta}-\frac{1}{m-\theta}\right). \qquad (4.18)\end{aligned}$$

Proof. (4.17) and (4.18) easily follow from (4.16):

- if $K>L$, $b-y<0\ \forall y\in[k,\infty]$. Then, using Lemma 4.4.4 and (4.16) yields

$$\widehat{PDIC}^{\star}(x,\lambda;K,L;r,\delta)=\frac{\psi(-\theta\sqrt{D})\,\mathrm{e}^{2b\theta}}{\theta\psi(\theta\sqrt{D})}\int_k^{\infty}\mathrm{e}^{(m-\theta)y}(x\,\mathrm{e}^{\sigma y}-K)dy,$$

  and the result easily follows.

- if $K<L$, $b-y$ is positive on $[k,b]$ and negative on $[b,\infty]$. We have to split the integral

$$I\triangleq\int_k^{+\infty}e^{my}(xe^{\sigma y}-K)K_{\lambda,D}(b-y)dy.$$

  appearing in (4.16).

$$I=\int_k^{b}e^{my}(xe^{\sigma y}-K)K_{\lambda,D}(b-y)dy+\int_b^{+\infty}e^{my}(xe^{\sigma y}-K)K_{\lambda,D}(b-y)dy\triangleq I_1+I_2. \qquad (4.19)$$

$$I_1 = D\psi(-\theta\sqrt{D})e^{\theta b}\int_b^{+\infty} e^{my}(xe^{\sigma y}-K)e^{-\theta y} = D\psi(-\theta\sqrt{D})e^{mb}\left(\frac{K}{m-\theta}-\frac{L}{m+\sigma-\theta}\right).$$

The integral I_2 can be split into three terms

$$\begin{aligned}
I_{21} &= D\psi(\theta\sqrt{D})\int_k^b e^{my}(xe^{\sigma y}-K)e^{\theta(y-b)}dy,\\
I_{22} &= -D\theta\sqrt{2\pi D}e^{\lambda D}\int_k^b e^{my}(xe^{\sigma y}-K)e^{\theta(y-b)}\mathcal{N}(\theta\sqrt{D}-\frac{b-y}{\sqrt{D}})dy,\\
I_{23} &= -D\theta\sqrt{2\pi D}e^{\lambda D}\int_k^b e^{my}(xe^{\sigma y}-K)e^{\theta(b-y)}\mathcal{N}(-\theta\sqrt{D}-\frac{b-y}{\sqrt{D}})dy.
\end{aligned}$$

An easy computation leads to

$$I_{21} = D\psi(\theta\sqrt{D})e^{-\theta b}\left\{e^{(m+\theta)b}\left[\frac{L}{m+\sigma+\theta}-\frac{K}{m+\theta}\right]+Ke^{(m+\theta)k}\left[\frac{1}{m+\theta}-\frac{1}{m+\sigma+\theta}\right]\right\}.$$

I_{22} and I_{23} are computed in the following way: we change variables (we introduce $v=\theta\sqrt{D}-\frac{b-y}{\sqrt{D}}$ (for the valuation of I_{22})) and we use the following equality $\int_{a_1}^{a_2}\mathcal{N}(v)e^{bv}dv = \frac{1}{b}[\mathcal{N}(a_2)e^{a_2 b}-\mathcal{N}(a_1)e^{a_1 b}-e^{\frac{b^2}{2}}(\mathcal{N}(a_2-b)-\mathcal{N}(a_1-b))]$, for $a_1, a_2, b\in\mathbb{R}$, $b\neq 0$ and $a_1\le a_2$.

We refer to Proposition 4.7.1, to prove that λ must be greater than $\frac{(m+\sigma)^2}{2}$. Let us prove (4.16). Using Proposition 4.4.1, we get

$$\widehat{PDIC}^{\star}(x,\lambda;K,L;r,\delta) = \int_k^\infty e^{my}(xe^{\sigma y}-K)\int_0^\infty \mathrm{e}^{-\lambda t}\, h_b^-(t,y)dtdy. \tag{4.20}$$

We would like to compute $\widehat{h_b^-}(\lambda,y) = \int_0^\infty \mathrm{e}^{-\lambda t}\, h_b^-(t,y)dt$. Using the definition of $h_b^-(t,y)$ in Proposition 4.4.1 yields

$$\widehat{h_b^-}(\lambda,y) = \int_{-\infty}^b \frac{b-z}{D}\,\mathrm{e}^{-\frac{(z-b)^2}{2D}}\int_0^\infty \mathrm{e}^{-\lambda t}\,\gamma^-(t,z-y)dt\,dz. \tag{4.21}$$

So, we need to compute the Laplace transform of $\gamma^-(t,x)$.

$$\int_0^\infty \mathrm{e}^{-\lambda t}\,\gamma^-(t,x)dt = \mathbb{E}_{\mathbb{P}}\left(\int_{T_b^-}^\infty \mathrm{e}^{-\lambda t}\,\frac{\mathrm{e}^{-\frac{x^2}{2(t-T_b^-)}}}{\sqrt{2\pi(t-T_b^-)}}dt\right).$$

The change of variables $u=t-T_b^-$ gives

$$\int_0^\infty \mathrm{e}^{-\lambda t}\,\gamma^-(t,x)dt = \mathbb{E}_{\mathbb{P}}(\mathrm{e}^{-\lambda T_b^-})\int_0^\infty \mathrm{e}^{-\lambda u}\,\frac{\mathrm{e}^{-\frac{x^2}{2u}}}{\sqrt{2\pi u}}du.$$

Using results from Appendices B.1 and B.2, we get

$$\int_0^\infty \mathrm{e}^{-\lambda t}\gamma^-(t,x)dt = \frac{\mathrm{e}^{-(|x|-b)\theta}}{\theta\psi(\theta\sqrt{D})}.$$

Thanks to (4.21), we can rewrite

$$\widehat{h_b^-}(\lambda, y) = \frac{\mathrm{e}^{b\theta}}{D\theta\psi(\theta\sqrt{D})}\int_{-\infty}^b (b-z)\,\mathrm{e}^{-\frac{(z-b)^2}{2D}-|z-y|\theta}\,dz.$$

By changing variables $v = b - z$, we obtain

$$\widehat{h_b^-}(\lambda, y) = \frac{\mathrm{e}^{b\theta}}{D\theta\psi(\theta\sqrt{D})}\int_0^\infty v\,\mathrm{e}^{-\frac{v^2}{2D}-|b-v-y|\theta}\,dv, \tag{4.22}$$

and (4.16) follows. ■

4.5 The Parisian Up calls

This section is devoted to the computation of the Laplace transforms of the Parisian Up call prices. We will go exactly through the same points as in the previous section but dealing with an Up and In call with $b \geq 0$ instead of a Down and In call with $b \leq 0$.

4.5.1 The valuation of a Parisian Up and In call with $b \geq 0$

The owner of an Up and In option receives the payoff if S makes an excursion above the level L longer than D before the maturity time T, which is exactly the same as saying that the Brownian motion Z makes an excursion above b longer than D. Using the previous notations, the price of a Parisian Up and In call is given by

$$PUIC^\star(x,T;K,L;r,\delta) = \mathbb{E}_{\mathbb{P}}(\mathbf{1}_{\{T_b^+<T\}}(x\,\mathrm{e}^{\sigma Z_T}-K)^+\,\mathrm{e}^{mZ_T}), \tag{4.23}$$

where T_b^+ is defined by (4.2). The valuation of $\widehat{PUIC}^\star$ in the case $b \geq 0$ is similar to the valuation of $\widehat{PDIC}^\star$ in the case $b \leq 0$ (see previous Section). Before computing the Laplace transform of $PUIC^\star$ in Theorem 4.5.2, we give a new expression for $PUIC^\star$.

Proposition 4.5.1.

$$PUIC^\star(x,T;K,L;r,\delta) = \int_k^\infty e^{my}(xe^{\sigma y}-K)h_b^+(T,y)dy,$$

where

$$h_b^+(t,y) = \int_b^\infty \frac{z-b}{D}\,\mathrm{e}^{-\frac{(z-b)^2}{2D}}\,\gamma^+(t,z-y)dz,$$

and

$$\gamma^{+}(t,x) = \mathbb{E}_{\mathbb{P}}\left(\mathbf{1}_{\{T_b^+ < t\}} \frac{\mathrm{e}^{-\frac{x^2}{2(t-T_b^+)}}}{\sqrt{2\pi(t-T_b^+)}} \right). \tag{4.24}$$

The proof of Proposition 4.5.1 is the same as the proof of Proposition 4.4.1. We only need to replace T_b^- by T_b^+.

Theorem 4.5.2. *The value of* $\widehat{PUIC}^{\star}$ *is given by the following formula*

$$\widehat{PUIC}^{\star}(x,\lambda;K,L;r,\delta) = \frac{\mathrm{e}^{-\theta b}}{D\theta\psi(\theta\sqrt{D})} \int_k^{\infty} e^{my}(xe^{\sigma y} - K) K_{\lambda,D}(y-b)dy. \tag{4.25}$$

For any $\lambda > \frac{(m+\sigma)^2}{2}$*, we get for* $K > L$

$$\begin{aligned}
\widehat{PUIC}^{\star}(x,\lambda;K,L;r,\delta) = 2\,\mathrm{e}^{(m-\theta)b}\, \frac{\sqrt{2\pi D}}{\psi(\theta\sqrt{D})} \Bigg[& \frac{K}{m^2-\theta^2}\,\mathrm{e}^{\frac{Dm^2}{2}}\, m\mathcal{N}(m\sqrt{D}+d) \\
& - \frac{L}{(m+\sigma)^2-\theta^2}\,\mathrm{e}^{\frac{D(m+\sigma)^2}{2}} (m+\sigma)\mathcal{N}((m+\sigma)\sqrt{D}+d) \Bigg] \\
+ \frac{\mathrm{e}^{-2b\theta}}{\psi(\theta\sqrt{D})} K\,\mathrm{e}^{(m+\theta)k}\,\mathrm{e}^{\lambda D}\,\sqrt{2\pi D}\mathcal{N}(d-\theta\sqrt{D}) & \left(\frac{1}{m+\sigma+\theta} - \frac{1}{m+\theta} \right) \\
+ \frac{\mathrm{e}^{(m-\theta)k}}{\theta\psi(\theta\sqrt{D})} K \left(\frac{1}{m-\theta} - \frac{1}{m+\sigma-\theta} \right) \Big(\psi(\theta\sqrt{D}) & - \theta\sqrt{2\pi D}\,\mathrm{e}^{\lambda D}\,\mathcal{N}(d+\theta\sqrt{D}) \Big).
\end{aligned} \tag{4.26}$$

and for $K \le L$

$$\begin{aligned}
\widehat{PUIC}^{\star}(x,\lambda;K,L;r,\delta) = & \frac{2\,\mathrm{e}^{(m-\theta)b}}{\psi(\theta\sqrt{D})} \left[\frac{K}{m^2-\theta^2}\psi(m\sqrt{D}) - \frac{L}{(m+\sigma)^2-\theta^2}\psi((m+\sigma)\sqrt{D}) \right] \\
& + \frac{\mathrm{e}^{-2b\theta}\,\psi(-\theta\sqrt{D})}{\theta\psi(\theta\sqrt{D})} K e^{(m+\theta)k} \left(\frac{1}{m+\theta} - \frac{1}{m+\theta+\sigma} \right).
\end{aligned} \tag{4.27}$$

Even if the computations involved in the proof of Theorem 4.5.2 are different from the one of Theorem 4.4.5, we dare omit the proof here as the scheme of the proof of Theorem 4.4.5 applies to the case of Up and In call.

Proof. (4.26) and (4.27) easily follow from (4.25):

- if $K > L$, $y - b > 0$ $\forall y \in [k, \infty[$. Then, using Lemma 4.4.4 and (4.25) yields

$$\widehat{PUIC}^{\star}(x,\lambda;K,L;r,\delta) = \frac{\mathrm{e}^{-\theta b}}{D\theta\psi(\theta\sqrt{D})} \int_k^{\infty} e^{my}(xe^{\sigma y} - K) K_{\lambda,D}(y-b)dy,$$

where

$$K_{\lambda,D}(y-b) = e^{-\theta(y-b)}D\psi(\theta\sqrt{D}) - D\theta\sqrt{2\pi D}e^{\lambda D}\left\{\mathcal{N}(\theta\sqrt{D}-\frac{y-b}{\sqrt{D}})e^{-\theta(y-b)}\right.$$
$$\left.+\mathcal{N}(-\theta\sqrt{D}-\frac{y-b}{\sqrt{D}})e^{\theta(y-b)}\right\}.$$

As for the valuation of I_2, page 104, we compute the three terms

$$\begin{aligned}
I_1 &= D\psi(\theta\sqrt{D})\int_k^\infty e^{my}(xe^{\sigma y}-K)e^{-\theta(y-b)}dy,\\
I_2 &= -D\theta\sqrt{2\pi D}e^{\lambda D}\int_k^\infty e^{my}(xe^{\sigma y}-K)e^{-\theta(y-b)}\mathcal{N}(\theta\sqrt{D}-\frac{y-b}{\sqrt{D}})dy,\\
I_3 &= -D\theta\sqrt{2\pi D}e^{\lambda D}\int_k^b e^{my}(xe^{\sigma y}-K)e^{\theta(y-b)}\mathcal{N}(-\theta\sqrt{D}-\frac{y-b}{\sqrt{D}})dy.
\end{aligned}$$

I_2 and I_3 are computed in the following way: we change variables (we introduce $v=\theta\sqrt{D}-\frac{y-b}{\sqrt{D}}$ (for the valuation of I_2)) and we use the following equality $\int_{-\infty}^a \mathcal{N}(v)e^{bv}dv = \frac{1}{b}[\mathcal{N}(a)e^{ab} - e^{\frac{b^2}{2}}(\mathcal{N}(a-b))]$, for $a,b\in\mathbb{R}$, $b\geq 0$.

- if $K<L$, $y-b$ is negative on $[k,b]$ and positive on $[b,\infty[$. We have to split the integral

$$I \triangleq \int_k^{+\infty} e^{my}(xe^{\sigma y}-K)K_{\lambda,D}(y-b)dy. \tag{4.28}$$

appearing in (4.25).

$$I = \int_k^b e^{my}(xe^{\sigma y}-K)K_{\lambda,D}(y-b)dy + \int_b^{+\infty} e^{my}(xe^{\sigma y}-K)K_{\lambda,D}(y-b)dy \triangleq I_1+I_2.$$

An easy computation gives us

$$\begin{aligned}
I_1 =& D\psi(-\theta\sqrt{D})e^{-\theta b}\int_k^b e^{my}(xe^{\sigma y}-K)e^{\theta y}dy = D\psi(-\theta\sqrt{D})\\
&\left[e^{-\theta b}Ke^{(m+\theta)k}\left(\frac{1}{m+\theta}-\frac{1}{m+\sigma+\theta}\right)+e^{mb}\left(\frac{L}{m+\sigma+\theta}-\frac{K}{m+\theta}\right)\right].
\end{aligned}$$

The integral I_2 can be split into three terms

$$\begin{aligned}
I_{21} &= D\psi(\theta\sqrt{D})e^{\theta b}\int_b^\infty e^{my}(xe^{\sigma y}-K)e^{-\theta y}dy,\\
I_{22} &= -D\theta\sqrt{2\pi D}e^{\lambda D}e^{\theta b}\int_b^\infty e^{my}(xe^{\sigma y}-K)e^{-\theta y}\mathcal{N}(\theta\sqrt{D}+\frac{b-y}{\sqrt{D}})dy,\\
I_{23} &= -D\theta\sqrt{2\pi D}e^{\lambda D}e^{-\theta b}\int_b^\infty e^{my}(xe^{\sigma y}-K)e^{\theta y}\mathcal{N}(-\theta\sqrt{D}+\frac{b-y}{\sqrt{D}})dy.
\end{aligned}$$

An easy computation leads to

$$I_{21} = D\psi(\theta\sqrt{D})e^{mb}\left(\frac{K}{m-\theta} - \frac{L}{m+\sigma-\theta}\right).$$

I_{22} and I_{23} are computed in the following way: we change variables (we introduce $v = \theta\sqrt{D} + \frac{b-y}{\sqrt{D}}$ (for the valuation of I_{22})) and we use the equality $\int_{a_1}^{a_2} \mathcal{N}(v)e^{bv}dv = \frac{1}{b}[\mathcal{N}(a_2)e^{a_2 b} - \mathcal{N}(a_1)e^{a_1 b} - e^{\frac{b^2}{2}}(\mathcal{N}(a_2-b) - \mathcal{N}(a_1-b))]$, for $a_1, a_2, b \in \mathbb{R}$, $b \neq 0$ and $a_1 \leq a_2$, as we did for the valuation of the PDIC, when $K < L$.

Let us prove (4.25). Using Proposition 4.5.1 yields

$$\widehat{PUIC}^{\star}(x, \lambda; K, L; r, \delta) = \int_k^{\infty} e^{my}(xe^{\sigma y} - K)\int_0^{\infty} \mathrm{e}^{-\lambda t}\, h_b^+(t, y)dtdy. \tag{4.29}$$

Following the proof of Theorem 4.4.5, we get

$$\widehat{h_b^+}(\lambda, y) = \int_b^{+\infty} \frac{z-b}{D}\,\mathrm{e}^{-\frac{(z-b)^2}{2D}} \int_0^{\infty} \mathrm{e}^{-\lambda t}\, \gamma_+(t, z-y)dt\, dz. \tag{4.30}$$

Using results from Appendices B.4 and B.2, we find

$$\int_0^{\infty} \mathrm{e}^{-\lambda t}\, \gamma_+(t, x)dt = \frac{\mathrm{e}^{-(|x|+b)\theta}}{\theta\psi(\theta\sqrt{D})}. \tag{4.31}$$

Thanks to (4.30), we can rewrite

$$\widehat{h_b^+}(\lambda, y) = \frac{\mathrm{e}^{-b\theta}}{D\theta\psi(\theta\sqrt{D})}\int_0^{\infty} x\,\mathrm{e}^{-\frac{x^2}{2D} - |b+x-y|\theta}\, dx. \tag{4.32}$$

and (4.25) follows. ■

4.6 Prices at any time t

So far, we have explained how to compute the prices at time 0 of the Laplace transforms of the different Parisian option prices w.r.t. maturity time. From a practical point of view, the real prize is to be able to hedge these options. This requires to compute the option prices at any given time t between 0 and the maturity time T. In this part, we explain how to deduce the prices at any time $t > 0$ from the prices at time 0.

In the following, we have chosen to focus on the Down and In call but the formula we obtain can easily be extended to the other options by means of parity relationships. We assume in the following computations that the relevant excursion has not occurred yet, otherwise the Parisian option has turned into the corresponding vanilla option and its price at time t is of common knowledge.

4.6.1 Down and In call

We introduce the r.v. D_t to count the time already spent in the excursion below b straddling time t

$$D_t = \begin{cases} t - g_t^b & \text{if } S_t \le b, \\ 0 & \text{if } S_t > b. \end{cases} \tag{4.33}$$

Note that D_t is $\mathcal{F}_t$-measurable.

Let $PDIC(t, D_t, S_t, T; K, L; r, \delta)$ be the price of a Down and In call at time t. We know that

$$PDIC(t, D_t, S_t, T; K, L; r, \delta) = \mathrm{e}^{-r(T-t)}\, \mathbb{E}_{\mathbb{Q}} \left((x\, \mathrm{e}^{\sigma(W_T + mT)} - K)^+ \, \mathbf{1}_{\{T_b^- \le T\}} | \mathcal{F}_t \right). \tag{4.34}$$

Proposition 4.6.1. *On the set $\{T_b^- > t\}$,*

$$\begin{aligned} & PDIC(t, D_t, S_t, T; K, L; r, \delta) \\ = \; & \mathrm{e}^{-\left(r + \frac{m^2}{2}\right)T'} \left\{ \mathbf{1}_{\{S_t > L\}} \mathbb{E} \left(\mathrm{e}^{mZ'_{T'}} (x\, \mathrm{e}^{\sigma Z'_{T'}} - K)^+ \, \mathbf{1}_{\{T'^-_{b'} \le T'\}} \right)_{|x = S_t} \right. \\ & + \mathbf{1}_{\{S_t \le L\}} \mathbf{1}_{\{D - D_t \le T'\}} \mathbb{E} \left(\mathrm{e}^{mZ'_{T'}} (x\, \mathrm{e}^{\sigma Z'_{T'}} - K)^+ \, \mathbf{1}_{\{T'_{b'} \ge D - d\}} \right)_{|x = S_t, d = D_t} \\ & \left. + \mathbf{1}_{\{S_t \le L\}} \mathbb{E} \left(\mathrm{e}^{mZ'_{T'}} (x\, \mathrm{e}^{\sigma Z'_{T'}} - K)^+ \, \mathbf{1}_{\{T'_{b'} \le D - d\}} \mathbf{1}_{\{T'^-_{b'} \le T'\}} \right)_{|x = S_t, d = D_t} \right\}. \end{aligned} \tag{4.35}$$

where Z' is a $\mathbb{P}-$Brownian motion independent of $\mathcal{F}_t$ and

$$T' = T - t, \quad b' = \frac{1}{\sigma} \ln \left(\frac{L}{x} \right), \quad T'^-_{b'} = T^-_{b'}(Z'), \quad T'_{b'} = T_{b'}(Z'). \tag{4.36}$$

Proof. We can change the probability measure as we did at the beginning to make $Z = \{W_t + mt; t \ge 0\}$ a $\mathbb{P}-$Brownian motion ($\mathbb{P}$ is defined in Proposition 4.2.1). $\mathbb{E}$ denotes the expectation under $\mathbb{P}$. Then, by changing the probability in Equation (4.34) we can write

$$\begin{aligned} & PDIC(t, D_t, S_t, T; K, L; r, \delta) \\ = \; & \mathrm{e}^{-r(T-t)} \frac{\mathbb{E} \left(\mathrm{e}^{mZ_T - \frac{1}{2}m^2 T} (x\, \mathrm{e}^{\sigma Z_T} - K)^+ \, \mathbf{1}_{\{T_b^- \le T\}} | \mathcal{F}_t \right)}{\mathrm{e}^{mZ_t - \frac{1}{2}m^2 t}}, \\ = \; & \mathrm{e}^{-r(T-t)} \frac{\mathbb{E} \left(\mathrm{e}^{mZ_t}\, \mathrm{e}^{m(Z_T - Z_t) - \frac{1}{2}m^2 T} (x\, \mathrm{e}^{\sigma Z_T} - K)^+ \, \mathbf{1}_{\{T_b^- \le T\}} | \mathcal{F}_t \right)}{\mathrm{e}^{mZ_t - \frac{1}{2}m^2 t}}, \\ = \; & \mathrm{e}^{-(r + \frac{m^2}{2})(T-t)}\, \mathbb{E} \left(\mathrm{e}^{m(Z_T - Z_t)} (S_t\, \mathrm{e}^{\sigma(Z_T - Z_t)} - K)^+ \, \mathbf{1}_{\{T_b^- \le T\}} | \mathcal{F}_t \right). \end{aligned}$$

We introduce $Z'_s = Z_{t+s} - Z_t$ for all $s \ge 0$. Z' is a $\mathbb{P}-$Brownian motion independent of $\mathcal{F}_t$.

$$PDIC(t, D_t, S_t, T; K, L; r, \delta) \;=\; \mathrm{e}^{-(r + \frac{m^2}{2})T'}\, \mathbb{E} \left(\mathrm{e}^{mZ'_{T'}} (S_t\, \mathrm{e}^{\sigma Z'_{T'}} - K)^+ \, \mathbf{1}_{\{T_b^- \le T\}} | \mathcal{F}_t \right).$$

The indicator $\mathbf{1}_{\{T_b^- \leq T\}}$ can be split up in several parts describing the different possible evolutions of Z' (see Figure 4.3). Either Z_t is not smaller than b and a whole excursion must be completed before T', or Z is already in an excursion below b. In the latter case, there are two possibilities corresponding to the two curves in Figure 4.3: either the current excursion will last longer than D (green curve), or Z will cross b before $D - D_t$ (blue curve) and a new excursion has to completed before T'. Then, on the set $\{T_b^- > t\}$, the indicator $\mathbf{1}_{\{T_b^- \leq T\}}$ can be rewritten as follows

$$\mathbf{1}_{\{T_b^- \leq T\}} = \mathbf{1}_{\{Z_t > b\}} \mathbf{1}_{\{T_{b'}'^- \leq T'\}} + \mathbf{1}_{\{Z_t \leq b\}} \left(\mathbf{1}_{\{T_{b'}' \geq D - D_t\}} \mathbf{1}_{\{D - D_t \leq T'\}} + \mathbf{1}_{\{T_{b'}' < D - D_t\}} \mathbf{1}_{\{T_{b'}'^- \leq T'\}} \right).$$

To find Equation (4.35), it is sufficient to notice that T_b', $T_b'^-$ and Z' are independent of $\mathcal{F}_t-$, whereas S_t and D_t are $\mathcal{F}_t-$measurable. ■

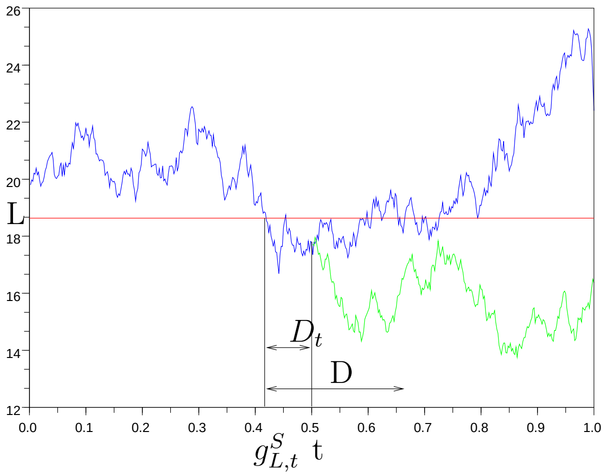

Figure 4.3: Possible evolutions of an asset price

In the sequel, we use the following decomposition based on Proposition 4.6.1

$$\begin{aligned} PDIC(t, D_t, S_t, T; K, L; r, \delta) \triangleq \mathrm{e}^{-\left(r + \frac{m^2}{2}\right)T'} \big\{ & \mathbf{1}_{\{S_t > L\}} E_1(S_t, T') \\ & + \mathbf{1}_{\{S_t \leq L\}} E_2(S_t, D_t, T') + \mathbf{1}_{\{S_t \leq L\}} E_3(S_t, D_t, T') \big\}. \end{aligned} \quad (4.37)$$

From Equation (4.35), we notice that E_1 is the star price of a Parisian Down and In call,

$$E_1(x, T') = PDIC^\star(x, T'; K, L; r, \delta). \quad (4.38)$$

Proposition 4.6.2. *On the set $\{T_b^- > t\}$, the price of a Down and In call at time t is given by*

$$\begin{aligned} PDIC^\star(t, D_t, S_t, T; K, L; r, \delta) = & \mathbf{1}_{\{Z_t > b\}} PDIC^\star(S_t, T - t, K, L; r, \delta) \\ & + \mathbf{1}_{\{Z_t \leq b\}} (\mathbf{1}_{\{D - D_t \leq T - t\}} BSC^\star(S_t, T - t; K; r, \delta) + g(S_t, D_t, T - t)) \end{aligned} \quad (4.39)$$

where the function g is characterized by its Laplace transform

$$\begin{aligned}\widehat{g}(S_t, D_t, \lambda) &= \mathrm{e}^{mb'} \int_0^{D-D_t} \mu_{b'}(u)\, \mathrm{e}^{-\lambda u}\, du \left(L\, \widehat{PDIC}^{\star,0}(\lambda; \frac{K}{L}; r, \delta) \right. \\ &\qquad \left. - \widehat{BSC}^{\star}(L, \lambda; K; r, \delta) \right).\end{aligned}$$

Proof. Let us go back to Equation (4.37). E_1 is already known (see (4.38)) and gives the first term on the r.h.s of (4.39). First, we deal with E_2 and after with E_3.

Step 1 : Laplace transform of E_2

$$\begin{aligned}E_2(x,d,t) &= \mathbb{E}\left(\mathrm{e}^{mZ'_t}(x\,\mathrm{e}^{\sigma Z'_t} - K)^+\, \mathbf{1}_{\{T'_{b'} \geq D-d\}}\right) \mathbf{1}_{\{D-d \leq t\}} \\ &= \mathbf{1}_{\{D-d\leq t\}} BSC^\star(x,t;K;r,\delta) - \mathbf{1}_{\{D-d\leq t\}}\mathbb{E}\left(\mathrm{e}^{mZ'_t}(x\,\mathrm{e}^{\sigma Z'_t} - K)^+\, \mathbf{1}_{\{T'_{b'} \leq D-d\}}\right) \\ &\triangleq E_{21}(x,d,t) - E_{22}(x,d,t).\end{aligned}$$

The term E_{21} corresponds to the first half of the second term on the r.h.s of (4.39). By conditioning w.r.t $\mathcal{F}_{T'_{b'}}$ and introducing $X_u = Z'_{u+T'_{b'}} - b'$, which is a Brownian motion independent of $\mathcal{F}_{T'_{b'}}$, we get

$$\begin{aligned}E_{22}(x,d,t) &= \mathbf{1}_{\{D-d\leq t\}}\mathbb{E}\left(\mathbf{1}_{\{T'_{b'}\leq D-d\}}\mathbb{E}\left(\mathrm{e}^{mX_{t-\tau}}\,\mathrm{e}^{mb'}(x\,\mathrm{e}^{\sigma b'}\,\mathrm{e}^{\sigma X_{t-\tau}} - K)^+\right)|_{\tau=T'_{b'}}\right), \\ &= \mathbf{1}_{\{D-d\leq t\}} \int_0^{D-d} \mathrm{e}^{mb'}\, \mathbb{E}(\mathrm{e}^{mX_{t-u}}(x\,\mathrm{e}^{\sigma b'}\,\mathrm{e}^{\sigma X_{t-u}} - K)^+)\mu_{b'}(u)du,\end{aligned}$$

where $\mu_{b'}$ is the density function of the hitting time $T'_{b'}$.

Using Proposition 4.2.7, it is quite easy to show that

$$\widehat{E_{22}}(S_t, D_t, \lambda) = \mathrm{e}^{mb'}\, \widehat{BSC^\star}(L, \lambda; K; r, \delta) \int_0^{D-D_t} \mathrm{e}^{-\lambda u}\, \mu_{b'}(u)du. \tag{4.40}$$

Step 2: Laplace transform of E_3 From Equation (4.37),

$$E_3(x,d,t) = \mathbb{E}\left(\mathrm{e}^{mZ'_t}(x\,\mathrm{e}^{\sigma Z'_t} - K)^+\, \mathbf{1}_{\{T'_{b'}\leq D-d\}}\mathbf{1}_{\{T'^-_{b'}\leq t\}}\right)$$

To compute E_3, we condition w.r.t $\mathcal{F}_{T'_{b'}}$ and introduce $X_u = Z'_{u+T'_{b'}} - b'$. X is a Brownian motion independent of $\mathcal{F}_{T'_{b'}}$. Hence, we get

$$\begin{aligned}E_3(x,d,t) &= \mathbb{E}\left(\mathbb{E}\left(\mathrm{e}^{mZ'_t}(x\,\mathrm{e}^{\sigma Z'_t} - K)^+\, \mathbf{1}_{\{T'_{b'}\leq D-d\}}\mathbf{1}_{\{T'^-_{b'}\leq t\}}|\mathcal{F}_{T'_{b'}}\right)\right), \\ &= \mathrm{e}^{mb'}\, \mathbb{E}\left(\mathbf{1}_{\{T'_{b'}\leq D-d\}}\mathbb{E}\left(\mathrm{e}^{mX_{t-T'_{b'}}}(x\,\mathrm{e}^{\sigma b'}\,\mathrm{e}^{\sigma X_{t-T'_{b'}}} - K)^+\, \mathbf{1}_{\{T'^-_{b'}\leq t\}}|\mathcal{F}_{T'_{b'}}\right)\right).\end{aligned}$$

Moreover on the set $\{T'_{b'} \leq D-d\}$, $T'^{-}_{b'}(Z') = T'_{b'}(Z') + T_0^-(X)$ a.s.. Hence, we find

$$\begin{aligned} E_3(x,d,t) &= \mathrm{e}^{mb'}\,\mathbb{E}\left(\mathbf{1}_{\{T'_{b'}\leq D-d\}}\mathbb{E}\left(\mathrm{e}^{mX_{t-T'_{b'}}}(x\,\mathrm{e}^{\sigma b'}\,\mathrm{e}^{\sigma X_{t-T'_{b'}}}-K)^+\,\mathbf{1}_{\{T_0^-\leq t-T'_{b'}\}}|\mathcal{F}_{T'_{b'}}\right)\right) \\ &= \mathrm{e}^{mb'}\,\mathbb{E}\left(\mathbf{1}_{\{\tau\leq D-d\}}\mathbb{E}\left(\mathrm{e}^{mX_{t-\tau}}(x\,\mathrm{e}^{\sigma b'}\,\mathrm{e}^{\sigma X_{t-\tau}}-K)^+\,\mathbf{1}_{\{T_0^-\leq t-\tau\}}\right)_{|\tau=T'_{b'}}\right) \\ &= \mathrm{e}^{mb'}\int_0^{D-d}\mathbb{E}\left(\mathrm{e}^{mX_{t-\tau}}(x\,\mathrm{e}^{\sigma b'}\,\mathrm{e}^{\sigma X_{t-\tau}}-K)^+\,\mathbf{1}_{\{T_0^-\leq t-\tau\}}\right)\mu_{b'}(\tau)d\tau. \end{aligned}$$

Using Proposition 4.2.7, one can show that

$$\widehat{E_3}(x,d,\lambda) = x\,\mathrm{e}^{(m+\sigma)b'}\,\widehat{PDIC}^{\star,0}(\lambda;K\,\mathrm{e}^{-\sigma b'}/x;r,\delta)\int_0^{D-d}\mathrm{e}^{-\lambda\tau}\,\mu_{b'}(\tau)d\tau.$$

Finally, we get

$$\widehat{E_3}(S_t,D_t,\lambda) = L\,\mathrm{e}^{mb'}\,\widehat{PDIC}^{\star,0}(\lambda;K/L;r,\delta)\int_0^{D-D_t}\mathrm{e}^{-\lambda\tau}\,\mu_{b'}(\tau)d\tau. \tag{4.41}$$

Noticing that $E_3(S_t,D_t,\lambda) - E_{22}(S_t,D_t,\lambda) = \widehat{g}(S_t,D_t,\lambda)$ ends the proof. ■

4.6.2 Other Parisian options

The price at time t of an Up and In call can be computed by closely following what has been done for the Down and In call and it is sufficient to replace $PDIC$ by $PUIC$ in the above formula. All the other Parisian option prices can be deduced using either an In and Out parity or a call put parity relationship.

4.7 The inversion of Laplace transforms

This section is devoted to the numerical inversion of the Laplace transforms computed previously. We recall that the Laplace transforms are computed with respect to the maturity time. We explain how to recover a function from its Laplace transform using a contour integral. The real problem is how to numerically evaluate this complex integral. This is done in two separate steps involving two different approximations. First, as explained in Section 5.5.1 we replace the integral by a series. The first step creates a discretisation error, which is handled by Proposition 5.5.2. Secondly, one has to compute a non-finite series. This can be achieved by simply truncating the series but it leads to a tremendously slow convergence. Here, we prefer to use the Euler acceleration as presented in Section 5.5.2. Proposition 5.5.3 states an upper-bound for the error due to the accelerated computation of the non finite series. Theorem 5.5.4 gives a bound for the global error.

4.7.1 Analytical prolongations

Because the Laplace inversion is performed in the complex plane, we have to extend to the complex plane the expressions obtained for the Laplace transforms computed above. To do so, we introduce the analytic prolongation of the normal cumulative distribution function on the complex plane. From Proposition 4.7.1, it is quite easy to show that the expressions obtained for a real value of the Laplace parameter are still valid for a complex one with the function $\mathcal{N}$ defined by Lemma 4.7.2.

Proposition 4.7.1 (abscissa of convergence)**.** *The abscissa of convergence of the Laplace transforms of the star prices of Parisian options is smaller than* $\frac{(m+\sigma)^2}{2}$. *All these Laplace transforms are analytic on the complex half plane* $\{z \in \mathbb{C} : \mathcal{R}\mathrm{e}\,(z) > \frac{(m+\sigma)^2}{2}\}$.

Proof. It is sufficient to notice that the star price of a Parisian option is bounded by $\mathbb{E}(\mathrm{e}^{mZ_T}(x\,\mathrm{e}^{\sigma W_T} + K))$.

$$\mathbb{E}(\mathrm{e}^{mZ_T}(x\,\mathrm{e}^{\sigma W_T} + K)) \leq K\,\mathrm{e}^{\frac{m^2}{2}T} + x\,\mathrm{e}^{\frac{(m+\sigma)^2}{2}T} = \mathcal{O}(\mathrm{e}^{\frac{(m+\sigma)^2}{2}T}).$$

Hence, Widder [58, Theorem 2.1] yields that the abscissa of convergence of the Laplace transforms of the star prices is smaller that $\frac{(m+\sigma)^2}{2}$. The second part of the proposition ensues from Widder [58, Theorem 5.a]. ∎

Lemma 4.7.2 (Analytical prolongation of $\mathcal{N}$)**.** *The unique analytic prolongation of the normal cumulative distribution function on the complex plane is defined by*

$$\mathcal{N}(x+iy) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} \mathrm{e}^{-\frac{(v+iy)^2}{2}}\, dv. \tag{4.42}$$

Proof. It is sufficient to notice that the function defined above is holomorphic on the complex plane (and hence analytic) and that it coincides with the normal cumulative distribution function on the real axis. ∎

With the definition of $\mathcal{N}$ given by Equation (4.42), it is clear that all the expressions obtained so far for the Laplace transforms are also valid for complex values of λ satisfying $\mathcal{R}\mathrm{e}\,(\lambda) > \frac{(m+\sigma)^2}{2}$ since their are analytic on the complex half plane $\{z \in \mathbb{C} : \mathcal{R}\mathrm{e}\,(z) > \frac{(m+\sigma)^2}{2}\}$.

4.7.2 The Fourier series representation

Thanks to Widder [58, Theorem 9.2], we know how to recover a function from its Laplace transform.

Theorem 4.7.3. *Let f be a continuous function defined on $\mathbb{R}^+$ and α a positive number. Assume that the function $f(t)\,\mathrm{e}^{-\alpha t}$ is integrable. Then, given the Laplace transform $\hat{f}$, f can be recovered from the contour integral*

$$f(t) = \frac{1}{2\pi i} \int_{\alpha - i\infty}^{\alpha + i\infty} \mathrm{e}^{st}\, \hat{f}(s) ds, \quad t > 0. \tag{4.43}$$

The variable α has to be chosen greater than the abscissa of convergence of $\hat{f}$. In our case, α must be chosen strictly greater than $(m+\sigma)^2/2$.

For any real valued function satisfying the hypotheses of Theorem 4.7.3, we introduce a trapezoidal discretisation of Equation (4.43) of step π/t.

$$f_{\pi/t}(t) = \frac{e^{\alpha t}}{2t}\widehat{f}(\alpha) + \frac{e^{\alpha t}}{t}\sum_{k=1}^{\infty}(-1)^k\,\mathcal{R}\mathrm{e}\left(\widehat{f}\left(\alpha + i\frac{k\pi}{t}\right)\right). \tag{4.44}$$

Proposition 4.7.4. *If f is a continuous bounded function satisfying $f(t)=0$ for $t<0$, we have*

$$\left|e_{\pi/t}(t)\right| \triangleq \left|f(t) - f_{\pi/t}(t)\right| \leq \|f\|_\infty \frac{e^{-2\alpha t}}{1-e^{-2\alpha t}}. \tag{4.45}$$

To prove Proposition 4.7.4,we need the following result adapted from Abate et al. [2, Theorem 5]

Lemma 4.7.5. *For any continuous and bounded function f such that $f(t)=0$ for $t<0$, we have*

$$e_{\pi/t}(t) = f_{\pi/t}(t) - f(t) = \sum_{\substack{k=-\infty \\ k\neq 0}}^{\infty} f\left(t(1+2k)\right)\mathrm{e}^{-2k\alpha t}. \tag{4.46}$$

Proof of Proposition 5.5.2. By performing a change of variables $s=\alpha+iu$ in the integral in (5.10), we can easily obtain an integral of a real variable.

$$f(t) = \frac{e^{\alpha t}}{2\pi}\int_{-\infty}^{+\infty}\widehat{f}(\alpha+iu)(\cos(ut)+i\sin(ut))du.$$

Moreover, since f is a real valued function, the imaginary part of the integral vanishes

$$f(t) = \frac{e^{\alpha t}}{2\pi}\int_{-\infty}^{+\infty}\mathcal{R}\mathrm{e}\left(\widehat{f}(\alpha+iu)\right)\cos(ut) - \mathcal{I}\mathrm{m}\left(\widehat{f}(\alpha+iu)\right)\sin(ut))du.$$

We notice that

$$\mathcal{I}\mathrm{m}\left(\widehat{f}(\alpha+iu)\right) = -\mathcal{I}\mathrm{m}\left(\widehat{f}(\alpha-iu)\right),\quad \mathcal{R}\mathrm{e}\left(\widehat{f}(\alpha+iu)\right) = \mathcal{R}\mathrm{e}\left(\widehat{f}(\alpha-iu)\right).$$

So,

$$f(t) = \frac{e^{\alpha t}}{\pi}\int_{0}^{+\infty}\mathcal{R}\mathrm{e}\left(\widehat{f}(\alpha+iu)\right)\cos(ut) - \mathcal{I}\mathrm{m}\left(\widehat{f}(\alpha+iu)\right)\sin(ut))du. \tag{4.47}$$

Using a trapezoidal integral with a step $h=\frac{\pi}{t}$ leads to Equation (5.11). Remembering that $f(t)=0$ for $t<0$, we can easily deduce from Lemma 4.7.5 that

$$e_{\pi/t}(t) = \sum_{k=1}^{\infty} f\left(t(1+2k)\right)\mathrm{e}^{-2k\alpha t}.$$

Taking the upper bound of f yields (5.12). ■

Remark 4.7.6. *For the upper bound in Proposition 5.5.2 to be smaller than* $10^{-8}\left\|f\right\|_\infty$*, one has to choose* $2\alpha t = 18.4$*. In fact, this bound holds for any choice of the discretisation step* h *satisfying* $h < 2\pi/t$*.*

Remark 4.7.7. *For the upper bound in Proposition 4.7.4 to be smaller than* $10^{-8}\left\|f\right\|_\infty$*, one has to choose* $2\alpha t = 18.4$*. In fact, this bound holds for any choice of the discretisation step* h *satisfying* $h < 2\pi/t$*.*

Simply truncating the series in the definition of $f_{\pi/t}$ to compute the trapezoidal integral is far too rough to provide a fast and accurate numerical inversion. One way to improve the convergence of the series is to use the Euler summation.

4.7.3 The Euler summation

To improve the convergence of a series S, we use the Euler summation technique as described by Abate et al. [2], which consists in computing the binomial average of q terms from the p-th term of the series S. The binomial average obviously converges to S as p goes to infinity. The following proposition describes the convergence rate of the binomial average to the infinite series $f_{\pi/t}(t)$ when p goes to ∞.

Proposition 4.7.8. *Let* f *be a function of class* $\mathcal{C}^{q+4}$ *such that there exists* $\epsilon > 0$ *s.t.* $\forall k \leq q+4,\ f^{(k)}(s) = \mathcal{O}(e^{(\alpha-\epsilon)s})$*, where* α *is the abscissa of convergence of* $\widehat{f}$*. We define* $s_p(t)$ *as the approximation of* $f_{\pi/t}(t)$ *when truncating the non-finite series in (4.44) to* p *terms*

$$s_p(t) = \frac{e^{\alpha t}}{2t}\widehat{f}(\alpha) + \frac{e^{\alpha t}}{t}\sum_{k=1}^{p}(-1)^k\,\mathcal{R}\mathrm{e}\left(\widehat{f}\left(\alpha + i\frac{\pi k}{t}\right)\right),$$

and $E(q,p,t) = \sum_{k=0}^{q} C_q^k 2^{-q} s_{p+k}(t)$*. Then,*

$$\left|f_{\pi/t}(t) - E(q,p,t)\right| \leq \frac{te^{\alpha t}\left|f'(0) - \alpha f(0)\right|}{\pi^2}\frac{p!\,(q+1)!}{2^q\,(p+q+2)!} + \mathcal{O}\left(\frac{1}{p^{q+3}}\right),$$

when p *goes to infinity.*

Using Propositions 4.7.4 and 4.7.8, we get the following result concerning the global error on the numerical computation of the price of a Parisian call option.

Corollary 4.7.9. *Let* f *be the price of a Parisian call option. Using the notations of Proposition 5.5.3, we have*

$$\left|f(t) - E(q,p,t)\right| \leq S_0\,\frac{e^{-2\alpha t}}{1-e^{-2\alpha t}} + \frac{e^{\alpha t}t\left|f'(0) - \alpha f(0)\right|p!\,(q+1)!}{\pi^2 2^q\,(p+q+2)!} + \mathcal{O}\left(\frac{1}{p^{q+3}}\right), \quad (4.48)$$

where α *is defined in Theorem 5.5.1.*

We refer the reader to Labart and Lelong [46] for a proof of Theorem 5.5.4 and Proposition 5.5.3.

For $2\alpha t = 18.4$ and $q = p = 15$, the global error is bounded by $S_0 10^{-8} + t\,|f'(0) - \alpha f(0)|\,10^{-11}$. As one can see, the method we use to invert Laplace transforms provides a very good accuracy with few computations.

Remark 4.7.10. *Considering the case of call options in Theorem 5.5.4 is sufficient since put prices are computed using parity relationships and their accuracy is hung up to the one of call prices.*

4.8 A few graphs

In this section, we perform a few numerical experiments with the method we have studied so far and compare it with the enhanced Monte Carlo method of Baldi et al. [10].

First, we consider a dynamic delta hedging simulation of a Parisian Up and Out call. We simulate an asset path and try to hedge along this trajectory. For this purpose, we use the formulae to derive the price of Parisian options at any time strictly positive. The delta simply ensues from a finite difference scheme. The discrete delta hedging proves quite efficient even though as one can see it on Figure 4.4, there are huge variations in the hedging portfolio when the option is about to be activated or canceled. This phenomena introduces some hedging error because the hedging is performed in discontinuous time. In this example, the hedging portfolio could be rebalanced three times a day.

Now, we would like to compare the prices obtained with our method with the prices given by the Monte Carlo method of Baldi et al. [10]. The Monte Carlo computation uses 10000 samples and 250 discretisation steps between 0 and T. Figure 4.5 shows the evolution w.r.t the delay of the price of a Down and Out put computed either with the invert Laplace transform method or the enhanced Monte Carlo method. The evolution of the prices provided by our method is much smoother than the one given by Monte Carlo. As one can see, the accuracy of the Monte Carlo method has nothing to do with the accuracy of our method. Let us recall that our prices are accurate up to 10^{-6} (when $S_0 = 100$) as stated in Theorem 5.5.4. Concerning the computational costs of the two methods, the invert Laplace transform method runs a thousand times faster than the corrected Monte Carlo.

4.9 Conclusion

In this work, we provide all the Laplace transforms of the different Parisian option prices, be it through explicit formulae or parity relationships. We also explain how to invert these formulae to compute the prices. The detailed study of the inversion algorithm enables to prove the accuracy and then the efficiency of the method. The

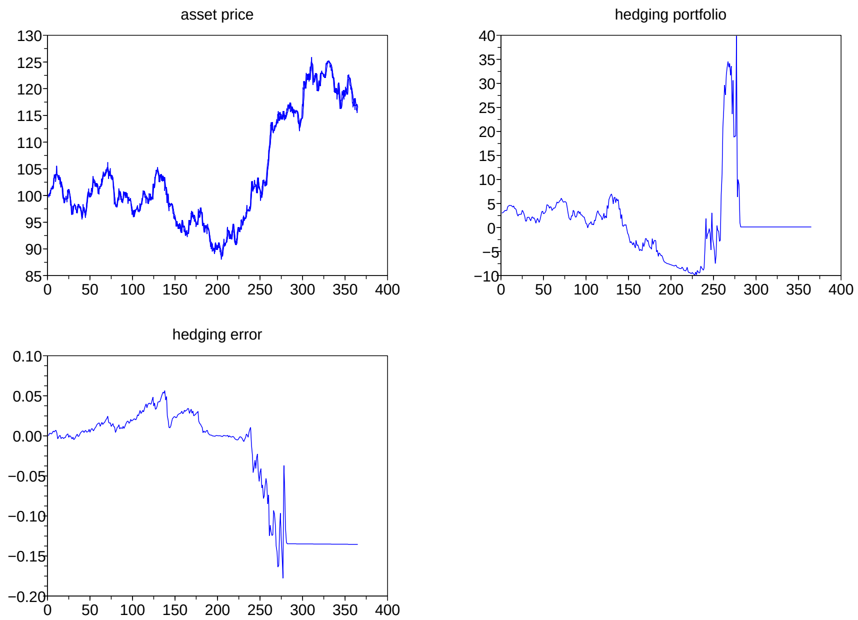

Figure 4.4: Example of delta hedging of a PUOC

$S_0 = 100 \quad K = 100 \quad T = 1 \quad L = 110$

$D = 20$ day $\quad \sigma = 0.2 \quad r = 0.025 \quad \delta = 0$

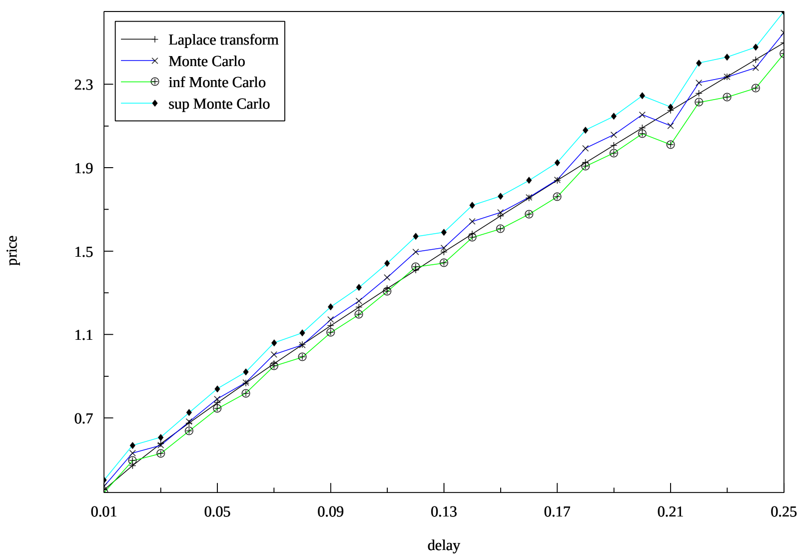

Figure 4.5: Comparison with improved Monte Carlo method in the case of a PDOP

$S_0 = 100 \quad K = 100 \quad T = 1 \quad L = 90$

$\sigma = 0.2 \quad r = 0.025 \quad \delta = 0$

efficiency is confirmed by the comparison with the enhanced Monte Carlo, which in fact is already very efficient when one thinks of how difficult it is to price Parisian options.

Chapter 5

Double barrier Parisian options

This chapter is based on an article Labart and Lelong [46] written with Céline LABART.

Abstract

In this work, we study a double barrier version of the standard Parisian options. We give closed formulae for the Laplace transforms of their prices with respect to the maturity time. We explain how to invert them numerically and prove a result on the accuracy of the numerical inversion.

5.1 Introduction

The pricing and hedging of vanilla options is now part of the common knowledge and the general interest has moved on to more complex products. So, practitioners need to be able to price these new products. Among them, there are the so-called path-dependent options. The ones we study in this paper are called double barrier Parisian options. They are a version with two barriers of the standard Parisian options introduced by Marc Chesney, Monique Jeanblanc and Marc Yor in 1997 (see Chesney et al. [23]). Before introducing double barrier Parisian options, we first recall the definition of Parisian options. Parisian options can be seen as barrier options where the condition involves the time spent in a row above or below a certain level, and not only an exiting time. Double barrier Parisian options are options where the conditions imposed on the asset involve the time spent out of the range defined by the two barriers.

The valuation of single barrier Parisian options can be done by using several different methods: Monte Carlo simulations, lattices, Laplace transforms or partial differential equations. As for standard barrier options, using simulations leads to a biased problem, due to the choice of the discretisation time step in the Monte Carlo algorithm. The problem of improving the performance of Monte Carlo methods in exotic pricing has drawn much attention and has particularly been developed by

Andersen and Brotherton-Ratcliffe [4]. Concerning lattices, we refer the reader to the work Avellaneda and Wu [8]. The idea of using Laplace transforms to price single barrier Parisian options is owed to Chesney et al. [23]. The Formulae of the Laplace transforms of all the different Parisian option prices can be found in Chapter 4. Schröder [57] and Hartley [39] have also studied these options using Laplace transforms. An approach based on partial differential equations has been developed by Haber et al. [36] and Wilmott [59]. Double Parisian options have already been priced by Baldi et al. [10] using Monte Carlo simulations corrected by the means of sharp large deviation estimates.

In this paper, we compute the prices of double barrier Parisian options by using Laplace transforms. First, we give a detailed computation of the Laplace transforms of the prices with respect to the maturity time. Then, we establish a formula for the inverse of the Laplace transforms using contour integrals. Since it cannot be computed exactly, we give an upper bound of the error between the approximated price and the exact one. We improve the approximation by using the Euler summation to get a fast and accurate numerical inversion. The paper is organised as follows. In section 5.2, we introduce the general framework and give precise definitions of double barrier Parisian option prices. In section 5.3, we establish a Call Put parity relationship, which enables us to deduce the price of put options from the prices of call options. In section 5.4, we carry out the computation of the Laplace transforms of double barrier Parisian option prices. In section 5.5, we give a formula for the inversion of the Laplace transforms and state some results concerning the accuracy of the method. The technique we use to prove these results is based on the regularity of option price (see Appendix 5.7). In section 5.6, we draw some graphs and compare the Laplace transform technique with the corrected Monte Carlo method of Baldi et al. [10]. For the comparison, we have used the implementation of the algorithm of Baldi et al. [10] available in PREMIA[1].

5.2 Definitions

5.2.1 Some notations

Let $S = \{S_t,\ t \geq 0\}$ denote the price of an underlying asset. We assume that under the risk neutral measure $\mathbb{Q}$, the dynamics of S is given by

$$dS_t = S_t((r-\delta)dt + \sigma dW_t),\ S_0 = x$$

where $W = \{W_t, t \geq 0\}$ is a $\mathbb{Q}-$Brownian motion, $x > 0$, the volatility σ is a positive constant, r denotes the interest rate. The parameter δ is the dividend rate if the underlying is a stock or the foreign interest rate in case of a currency. We assume that

[1]PREMIA is a pricing software developed the MathFi team of INRIA Rocquencourt, see `http://www.premia.fr`.

both r and δ are constant. It follows that

$$S_t = x\,\mathrm{e}^{(r-\delta-\sigma^2/2)t+\sigma W_t}\,.$$

We introduce

$$m = \frac{1}{\sigma}\left(r-\delta-\frac{\sigma^2}{2}\right). \tag{5.1}$$

Under $\mathbb{Q}$, the dynamics of the asset is given by $S_t = x\,\mathrm{e}^{\sigma(mt+W_t)}$. From now on, we consider that every option has a finite maturity time T. Relying on Girsanov's Theorem (see Revuz and Yor [54]), we can introduce a new probability $\mathbb{P}$ — defined by $\frac{d\mathbb{P}}{d\mathbb{Q}}_{|\mathcal{F}_T} = e^{mW_T-\frac{m^2}{2}T}$ — which makes $Z = \{Z_t = W_t + mt, 0 \le t \le T\}$ a $\mathbb{P}$-Brownian motion. Thus, S rewrites $S_t = x\,e^{\sigma Z_t}$ under $\mathbb{P}$. Without any further indications, all the processes and expectations are considered under $\mathbb{P}$.

5.2.2 Double barrier Parisian option

There are two different ways of measuring the time spent above or below a barrier. Either, one only counts the time spent in a row and resets the counting each time the stock price crosses the barrier(s) — we call it the *continuous* manner — or one adds the time spent in the relevant excursions without resuming the counting from 0 each time the stock price crosses the barrier(s) — we call it the *cumulative* manner. In practice, these two ways of counting time raise different questions about the paths of Brownian motion. In this work, we only focus on continuous style options.

Knock Out

A knock out double barrier Parisian call (respectively put) is lost if S makes an excursion outside the range (L_1, L_2) older than D before T otherwise it pays at maturity time T $(S_T - K)_+$ (respectively $(K - S_T)_+$) where K is the strike.

We introduce b_1 and b_2 the barriers corresponding to L_1 and L_2 for the Brownian motion Z

$$b_1 = \frac{1}{\sigma}\log\left(\frac{L_1}{x}\right), \quad b_2 = \frac{1}{\sigma}\log\left(\frac{L_2}{x}\right).$$

For some level b, let us introduce the following notations

$$\begin{aligned}
g_t^b = g_t^b(Z) &= \sup\{u \le t \mid Z_u = b\},\\
T_b^- = T_b^-(Z) &= \inf\{t > 0 \mid (t - g_t^b)\,\mathbf{1}_{\{Z_t<b\}} > D\},\\
T_b^+ = T_b^+(Z) &= \inf\{t > 0 \mid (t - g_t^b)\,\mathbf{1}_{\{Z_t>b\}} > D\}.
\end{aligned}$$

Hence, the price of a knock out double barrier Parisian call (DPOC) is given by

$$DPOC(x, T; K, L_1, L_2; r, \delta) = e^{-(\frac{m^2}{2}+r)T}\mathbb{E}\left[e^{mZ_T}(S_T - K)_+\mathbf{1}_{\{T_{b_1}^- > T\}}\mathbf{1}_{\{T_{b_2}^+ > T\}}\right]. \tag{5.2}$$

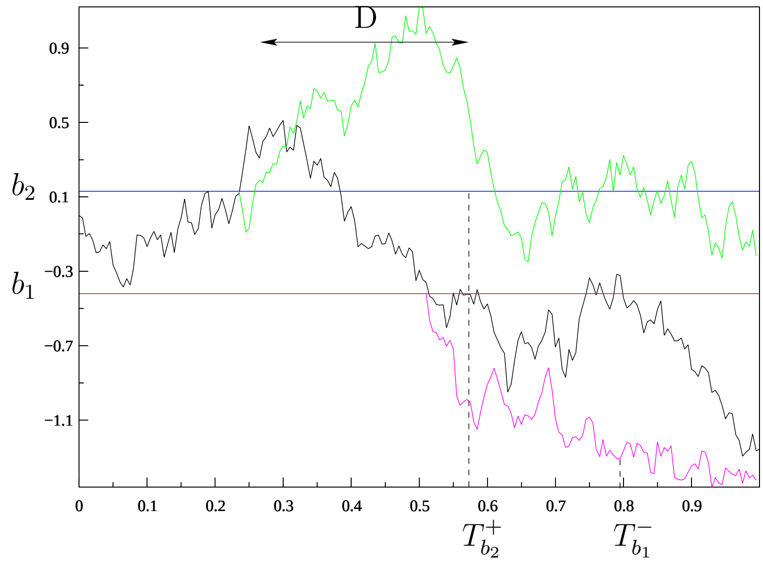

Figure 5.1: Brownian paths

The two indicators can be rewritten

$$\mathbf{1}_{\{T^-_{b_1}>T\}}\mathbf{1}_{\{T^+_{b_2}>T\}} = 1 - \mathbf{1}_{\{T^-_{b_1}<T\}} - \mathbf{1}_{\{T^+_{b_2}<T\}} + \mathbf{1}_{\{T^-_{b_1}<T\}}\mathbf{1}_{\{T^+_{b_2}<T\}}.$$

Since the random variables T^+_b and T^-_b have a density w.r.t the Lebesgue measure (see Appendix 5.8), one can use either strict or non-strict inequalities in the previous formula.

Dealing with inequalities of the type $\mathbf{1}_{\{T^\pm_b<T\}}$ is much simpler than $\mathbf{1}_{\{T^\pm_b>T\}}$since we can condition w.r.t. $\mathcal{F}_{T^\pm_b}$ and use the Strong Markov property. Consequently, Equation (5.2) can be split into four terms using the prices of single barrier Parisian options. To describe single barrier Parisian options, we use the following notations : PDOC means Parisian Down and Out Call, whereas PUIP stands for Parisian Up and In Put and so on. BSC simply denotes the price of a standard call option.

$$\begin{aligned} DPOC(x,T;K,L_1,L_2;r,\delta) =& BSC(x,T;K;r,\delta) - PDIC(x,T;K,L_1;r,\delta) \\ & - PUIC(x,T;K,L_2;r,\delta) + e^{-(\frac{m^2}{2}+r)T}A, \end{aligned} \tag{5.3}$$

where

$$A = \mathbb{E}\left[e^{mZ_T}(S_T-K)_+\mathbf{1}_{\{T^-_{b_1}<T\}}\mathbf{1}_{\{T^+_{b_2}<T\}}\right], \tag{5.4}$$

For any function f of the maturity T, we introduce the “star” notation

$$^\star f(T) = e^{(r+\frac{1}{2}m^2)T}f(T).$$

The computation of $^\star DPOC$ will be done using numerical inversion of its Laplace transform with respect to T. Explicit formulae for the Laplace transforms of the first three terms in (5.3) — $\widehat{^\star BSC}$, $\widehat{^\star PDIC}$, $\widehat{^\star PUIC}$ — can be found in Chapter 4 (see Labart and Lelong [45]). We only need to compute $\widehat{A} = \int_0^\infty \mathbb{E}\left[e^{mZ_u}(S_u - K)_+ \mathbf{1}_{\{T^-_{b_1}<u\}}\mathbf{1}_{\{T^+_{b_2}<u\}}\right] e^{-\lambda u} du$. A detailed computation can be found in Section 5.4.

Knock In

A knock in double barrier Parisian call (respectively put) pays at maturity time T $(S_T - K)_+$ (respectively $(K - S_T)_+$) if S makes an excursion outside the range (L_1, L_2) longer than D before T and is lost otherwise.

The price of such an option (DPIC) is given by

$$DPIC(x, T; K, L_1, L_2; r, \delta) = e^{-(\frac{m^2}{2}+r)T} \\ \mathbb{E}\left[e^{mZ_T}(S_T - K)_+ \left(\mathbf{1}_{\{T^-_{b_1}<T\}} + \mathbf{1}_{\{T^+_{b_2}<T\}} - \mathbf{1}_{\{T^-_{b_1}<T\}}\mathbf{1}_{\{T^+_{b_2}<T\}}\right)\right].$$

It is quite obvious that $DPIC$ can be expressed in terms of single barrier Parisian option prices and the quantity A (see (5.4))

$$DPIC(x, T; K, L_1, L_2; r, \delta) = PDIC(x, T; K, L_1; r, \delta) \\ + PUIC(x, T; K, L_2; r, \delta) - e^{-(\frac{m^2}{2}+r)T} A. \quad (5.5)$$

5.3 A Call Put relationship

As for single barrier Parisian options, a parity relationship between calls and puts holds. The basic idea of the relationship is that the processes Z and $-Z$ have the same law.

Therefore, introducing the new Brownian motion $\tilde{Z} = -Z$ enables to rewrite the price of double barrier Parisian puts

$$DPOP(x, T; K, L_1, L_2, D, r, \delta) = Kxe^{-\left(r+\frac{m^2}{2}\right)T} \\ \mathbb{E}\left(e^{-(m+\sigma)\tilde{Z}_T}\left(\frac{1}{x}e^{\sigma\tilde{Z}_T} - \frac{1}{K}\right)^+ \mathbf{1}_{\{T^+_{-b_2}>T\}}\mathbf{1}_{\{T^-_{-b_1}>T\}}\right). \quad (5.6)$$

Let us introduce

$$m' = -(m+\sigma), \quad \delta' = r, \quad r' = \delta, \quad b'_1 = -b_2, \quad b'_2 = -b_1.$$

One can easily check that $m' = \frac{1}{\sigma}\left(r' - \delta' - \frac{\sigma^2}{2}\right)$ and that $r' + \frac{m'^2}{2} = r + \frac{m^2}{2}$. Moreover, by noticing that the barrier L' corresponding to $b' = -b$ is $\frac{1}{L}$, it becomes clear that the expectation on the right hand side of (5.6) can be interpreted as

$$xK\ DPOC\left(\frac{1}{x}, T; \frac{1}{K}, \frac{1}{L_2}, \frac{1}{L_1}, D, \delta, r\right).$$

The same kind of relation holds for knock in options

$$DPIP(x,T;K,L_1,L_2,D,r,\delta) = xK\ DPIC\left(\frac{1}{x},T;\frac{1}{K},\frac{1}{L_2},\frac{1}{L_1},D,\delta,r\right).$$

5.4 Computation of Laplace transforms

The computation of $DPOC^\star$ and $DPIC^\star$ will be done using the numerical inversion of their Laplace transforms w.r.t. the maturity time. As explained in Equations (5.3) and (5.5), the computation of the Laplace transforms of $DPOC^\star$ and $DPIC^\star$ boils down to the one of A.

Theorem 5.4.1 (Laplace transform of Double barrier Parisian options)**.** *The following relations hold*

$$\begin{aligned}\widehat{DPOC}^\star(x,\lambda;K,L_1,L_2;r,\delta) =& \widehat{BSC}^\star(x,\lambda;K;r,\delta) - \widehat{PDIC}^\star(x,\lambda;K,L_1;r,\delta)\\ &- \widehat{PUIC}^\star(x,T;K,L_2;r,\delta) + \widehat{A},\\ \widehat{DPIC}^\star(x,T;K,L_1,L_2;r,\delta) =& \widehat{PDIC}^\star(x,T;K,L_1;r,\delta)\\ &+ \widehat{PUIC}^\star(x,T;K,L_2;r,\delta) - \widehat{A},\end{aligned}$$

where $\widehat{A}$ is the Laplace transform of A w.r.t. maturity time given by

$$\begin{aligned}\widehat{A} =& \mathbb{E}\left[e^{-\lambda T^-_{b_1}}\mathbf{1}_{\{T^-_{b_1}<T^+_{b_2}\}}\right]\mathbb{E}\left[e^{\sqrt{2\lambda}Z_{T^-_{b_1}}}\right]\widehat{PUIC}^\star_{|x<L_2}(x,\lambda;K,L_2;r,\delta)\\ &+\mathbb{E}\left[e^{-\lambda T^+_{b_2}}\mathbf{1}_{\{T^+_{b_2}<T^-_{b_1}\}}\right]\mathbb{E}\left[e^{-\sqrt{2\lambda}Z_{T^+_{b_2}}}\right]\widehat{PDIC}^\star_{|x>L_1}(x,\lambda;K,L_1;r,\delta),\end{aligned}$$

where $\widehat{PUIC}^\star_{|x<L_2}$ (resp. $\widehat{PDIC}^\star_{|x>L_1}$) means that we use the definition of $\widehat{PUIC}^\star$ (resp. $\widehat{PUIC}^\star$) in the case $x < L_2$ (resp. in the case $x > L_1$).

We refer to Appendix B.5 for explicit formulae for $\mathbb{E}[e^{-\lambda T^-_{b_1}}\mathbf{1}_{\{T^-_{b_1}<T^+_{b_2}\}}]$, $\mathbb{E}[e^{-\lambda T^+_{b_2}}\mathbf{1}_{\{T^+_{b_2}<T^-_{b_1}\}}]$, $\mathbb{E}[\exp(\sqrt{2\lambda}Z_{T^-_{b_1}})]$ and $\mathbb{E}[\exp(-\sqrt{2\lambda}Z_{T^+_{b_2}})]$.

Proof. The first part of the theorem ensues directly from Equations (5.3) and (5.5).

We split Equation (5.4) into two terms by introducing $\mathbf{1}_{\{T^-_{b_1}<T^+_{b_2}\}}$ and $\mathbf{1}_{\{T^+_{b_2}<T^-_{b_1}\}}$. This yields to

$$\begin{aligned}A =& \mathbb{E}\left[\mathbf{1}_{\{T^-_{b_1}<T\}}\mathbf{1}_{\{T^-_{b_1}<T^+_{b_2}\}}\mathbb{E}\left[e^{mZ_T}(x\,\mathrm{e}^{\sigma Z_T}-K)_+\mathbf{1}_{\{T^+_{b_2}<T\}}\middle|\mathcal{F}_{T^-_{b_1}}\right]\right]\\ &+\mathbb{E}\left[\mathbf{1}_{\{T^+_{b_2}<T\}}\mathbf{1}_{\{T^+_{b_2}<T^-_{b_1}\}}\mathbb{E}\left[e^{mZ_T}(x\,\mathrm{e}^{\sigma Z_T}-K)_+\mathbf{1}_{\{T^-_{b_1}<T\}}\middle|\mathcal{F}_{T^+_{b_2}}\right]\right] \triangleq A_1 + A_2.\end{aligned}$$

We concentrate on $\hat{A}_1$ as A_2 can be treated the same way. We aim at proving that

$$\widehat{A_1} = \mathbb{E}\left[e^{-\lambda T_{b_1}^-}\mathbf{1}_{\{T_{b_1}^- < T_{b_2}^+\}}\right]\ \mathbb{E}\left[\mathrm{e}^{\sqrt{2\lambda}Z_{T_{b_1}^-}}\right]\ \widehat{PUIC}^\star_{|x<L_2}(x,\lambda;K,L_2;r,\delta). \tag{5.7}$$

Equality (5.7) ensues from the two following relations we prove below

$$\widehat{A_1} = \mathbb{E}\left[e^{-\lambda T_{b_1}^-}\mathbf{1}_{\{T_{b_1}^- < T_{b_2}^+\}}\right]\ \mathbb{E}\left[\mathrm{e}^{mZ_{T_{b_1}^-}}\ \widehat{PUIC}^\star(x\,\mathrm{e}^{\sigma Z_{T_{b_1}^-}},\lambda;K,L_2;r,\delta)\right], \tag{5.8}$$

$$\mathbb{E}\left[\mathrm{e}^{mZ_{T_{b_1}^-}}\ \widehat{PUIC}^\star(x\,\mathrm{e}^{\sigma Z_{T_{b_1}^-}},\lambda;K,L_2;r,\delta)\right] = \mathbb{E}\left[\mathrm{e}^{\sqrt{2\lambda}Z_{T_{b_1}^-}}\right]\widehat{PUIC}^\star_{|x<L_2}(x,\lambda;K,L_2;r,\delta). \tag{5.9}$$

Step 1: Proof of (5.8).

$$\widehat{A_1} = \mathbb{E}\left[\int_0^\infty e^{-\lambda u}\mathbf{1}_{\{T_{b_1}^- < u\}}\mathbf{1}_{\{T_{b_1}^- < T_{b_2}^+\}}e^{mZ_{T_{b_1}^-}}PUIC^\star\left(xe^{\sigma Z_{T_{b_1}^-}}, u - T_{b_1}^-;K,L_2;r,\delta\right)du\right].$$

As a function of T, A_1 shows up as a convolution, hence its Laplace tranforms is given by

$$\widehat{A_1} = \mathbb{E}\left[e^{-\lambda T_{b_1}^-}\mathbf{1}_{\{T_{b_1}^- < T_{b_2}^+\}}e^{mZ_{T_{b_1}^-}}\widehat{PUIC}^\star(xe^{\sigma Z_{T_{b_1}^-}},\lambda;K,L_2;r,\delta)\right].$$

From Chesney et al. [23, Sections 8.3 and 8.4], we know that $T_{b_1}^-$ is an $\mathcal{F}_{g_t}^+$−stopping time whereas $Z_{T_{b_1}^-}$ is independent of $\mathcal{F}^+_{g_{T_{b_1}^-}}$. Hence, Step 1 is completed.

Step 2: Proof of (5.9).

Since $x\exp(\sigma Z_{T_{b_1}^-}) < L_2$ (as $Z_{T_{b_1}^-} < b_1$), we compute $\widehat{PUIC}^\star(x\exp(\sigma Z_{T_{b_1}^-}),\lambda;K,L_2;r,\delta)$ using the definition of $\widehat{PUIC}^\star(x,\lambda;K,L_2;r,\delta)$ in the case $x < L_2$. In such a case, we can write $\widehat{PUIC}^\star(x,\lambda;K,L_2;r,\delta) = x^{\frac{\sqrt{2\lambda}-m}{\sigma}}f(\lambda;K,L_2;r,\delta)$, where f can easily be deduced from the formulae recalled in Chapter 4

$$\mathbb{E}\left[e^{mZ_{T_{b_1}^-}}\widehat{PUIC}^\star(xe^{\sigma Z_{T_{b_1}^-}},\lambda;K,L_2;r,\delta)\right] = \mathbb{E}\left[e^{mZ_{T_{b_1}^-}}\left(x\,\mathrm{e}^{\sigma Z_{T_{b_1}^-}}\right)^{\frac{\sqrt{2\lambda}-m}{\sigma}}f(\lambda;K,L_2;r,\delta)\right],$$

and the result follows. ■

5.5 The inversion of Laplace transforms

This section is devoted to the numerical inversion of the Laplace transforms computed previously. We recall that the Laplace transforms are computed with respect to the maturity time. We explain how to recover a function from its Laplace transform using a contour integral. The real problem is how to numerically evaluate this complex integral. This is done in two separate steps involving two different errors. First, as explained in Section 5.5.1 we replace the integral by a series. The first step creates a discretisation error, which is handled by Proposition 5.5.2. Secondly, one has to compute a non-finite series. This can be achieved by simply truncating the series but it leads to a tremendously slow convergence. Here, we prefer to use the Euler acceleration as presented in Section 5.5.2. Proposition 5.5.3 states an upper-bound for the error due to the accelerated computation of the non finite series. Theorem 5.5.4 gives a bound for the global error.

5.5.1 The Fourier series representation

Thanks to Widder [58, Theorem 9.2], we know how to recover a function from its Laplace transform.

Theorem 5.5.1. *Let f be a continuous function defined on $\mathbb{R}^+$ and α a positive number. If the function $f(t)\,\mathrm{e}^{-\alpha t}$ is integrable, then given the Laplace transform $\hat{f}$, f can be recovered from the contour integral*

$$f(t) = \frac{1}{2\pi i}\int_{\alpha-i\infty}^{\alpha+i\infty} \mathrm{e}^{st}\,\hat{f}(s)ds, \quad t > 0. \tag{5.10}$$

The variable α has to be chosen greater than the abscissa of convergence of $\hat{f}$. The abscissa of convergence of the Laplace transforms of the double barrier Parisian option prices computed previously is smaller than $(m+\sigma)^2/2$. Hence, α must be chosen strictly greater than $(m+\sigma)^2/2$.

For any real valued function satisfying the hypotheses of Theorem 5.5.1, we introduce a trapezoidal discretisation of Equation (5.10)

$$f_{\pi/t}(t) = \frac{e^{\alpha t}}{2t}\widehat{f}(\alpha) + \frac{e^{\alpha t}}{t}\sum_{k=1}^{\infty}(-1)^k\,\mathcal{R}\mathrm{e}\left(\widehat{f}\left(\alpha + i\frac{k\pi}{t}\right)\right). \tag{5.11}$$

We recall here a result from Chapter 4 (Proposition 4.7.4)

Proposition 5.5.2. *If f is a continuous bounded function satisfying $f(t) = 0$ for $t < 0$, we have*

$$\left|e_{\pi/t}(t)\right| = \left|f(t) - f_{\pi/t}(t)\right| \le \|f\|_\infty\,\frac{e^{-2\alpha t}}{1-e^{-2\alpha t}}. \tag{5.12}$$

Simply truncating the summation in the definition of $f_{\pi/t}$ to compute the trapezoidal integral is far too rough to provide a fast and accurate numerical inversion. One way to improve the convergence of the series is to use the Euler summation.

5.5.2 The Euler summation

To improve the convergence of a series S, we use the Euler summation technique as described by Abate et al. [2], which consists in computing the binomial average of q terms from the p-th term of the series S. The binomial average obviously converges to S as p goes to infinity. The following proposition describes the convergence rate of the binomial average to the infinite series $f_{\pi/t}(t)$ when p goes to ∞.

Proposition 5.5.3. *Let f be a function of class $\mathcal{C}^{q+4}$ such that there exists $\epsilon > 0$ s.t. $\forall k \leq q+4$, $f^{(k)}(s) = \mathcal{O}(e^{(\alpha-\epsilon)s})$. We define $s_p(t)$ as the approximation of $f_{\pi/t}(t)$ when truncating the non-finite series in (5.11) to p terms*

$$s_p(t) = \frac{e^{\alpha t}}{2t}\widehat{f}(\alpha) + \frac{e^{\alpha t}}{t}\sum_{k=1}^{p}(-1)^k\,\mathcal{R}\mathrm{e}\left(\widehat{f}\left(\alpha + i\frac{\pi k}{t}\right)\right), \tag{5.13}$$

and $E(q,p,t) = \sum_{k=0}^{q} C_q^k 2^{-q} s_{p+k}(t)$. Then,

$$\left|f_{\pi/t}(t) - E(q,p,t)\right| \leq \frac{te^{\alpha t}\,|f'(0) - \alpha f(0)|}{\pi^2}\frac{p!\,(q+1)!}{2^q\,(p+q+2)!} + \mathcal{O}\left(\frac{1}{p^{q+3}}\right)$$

when p goes to infinity.

Using Propositions 5.5.2 and 5.5.3, we get the following result concerning the global error on the numerical computation of the price of a double barrier Parisian call option

Theorem 5.5.4. *Let f be the price of a double barrier Parisian call option. Using the notations of Proposition 5.5.3, we have*

$$|f(t) - E(q,p,t)| \leq S_0\,\frac{e^{-2\alpha t}}{1-e^{-2\alpha t}} + \frac{e^{\alpha t}t\,|f'(0) - \alpha f(0)|\,p!\,(q+1)!}{\pi^2 2^q\,(p+q+2)!} + \mathcal{O}\left(\frac{1}{p^{q+3}}\right) \tag{5.14}$$

where α is defined in Theorem 5.5.1.

Proof of Theorem 5.5.4. f being the price of a double barrier Parisian call option, we know that f is bounded by S_0. Moreover, f is continuous (actually of class $\mathcal{C}^\infty$, see Appendix 5.7). Hence, Proposition 5.5.2 yields the first term on the right-hand side of (5.14).

Relying on Proposition 5.7.1, we know that ${}^\star f$ is of class $\mathcal{C}^\infty$ and ${}^\star f^{(k)}(t) = \mathcal{O}(e^{\frac{(m+\sigma)^2}{2}t})$, $\forall k \geq 0$. Since $f(t) = \mathrm{e}^{-(r+m^2/2)t}\ {}^\star f(t)$, it is quite obvious that f is also of class $\mathcal{C}^\infty$ and $f^{(k)}(t) = \mathcal{O}\left(\mathrm{e}^{\left((m+\sigma)^2/2-(r+m^2/2)\right)t}\right)$, $\forall k \geq 0$. Since $\alpha > \frac{(m+\sigma)^2}{2}$, we can apply Proposition 5.5.3 to get the result. ■

Proof of Proposition 5.5.3. We compute the difference between two successive terms.

$$E(q,p+1,t) - E(q,p,t) = \frac{\mathrm{e}^{\alpha t}}{2^q t}\sum_{k=0}^{q} C_q^k(-1)^{p+1+k}a_{p+k+1},$$

where

$$a_p = \int_0^{+\infty} e^{-\alpha s} \cos\left(\frac{p}{t}\pi s\right) f(s) ds. \tag{5.15}$$

Let $g(s) = e^{-\alpha s} f(s)$. Since $g^{(k)}(\infty) = 0$ for $k \leq q+3$ and $g^{(q+4)}$ is integrable, we can perform $(q+3)$ integrations by parts in (5.15) to obtain a Taylor expansion when p goes to infinity

$$a_p = \frac{c_2}{p^2} + \frac{c_4}{p^4} + \cdots + \frac{c_q}{p^{2[(q+3)/2]}} + \mathcal{O}\left(\frac{1}{p^{q+4}}\right) \tag{5.16}$$

with $c_2 = \frac{4t^2(f'(0)-\alpha f(0))}{\pi^2}$.

We can rewrite (5.16)

$$a_p = \frac{c_2}{p(p+1)} + \frac{c_3'}{p(p+1)(p+2)} + \cdots + \frac{c_q'}{p(p+1)\cdots(p+q+2)} + \mathcal{O}\left(\frac{1}{p^{q+4}}\right).$$

Some elementary computations show that for $j \geq 2$

$$\sum_{k=0}^{q} C_q^k (-1)^{p+1+k} \frac{1}{(p+k)(p+k+1)\cdots(p+k+j)} = (-1)^{p+1} \frac{p!\,(q+j-1)!}{(j-1)!(p+q+j)!}.$$

Computing $\sum_{k=0}^{q} C_q^k (-1)^{p+1+k} a_{p+k+1}$ leads to

$$E(q,p+1,t) - E(q,p,t) = (-1)^{p+1}\frac{\mathrm{e}^{\alpha t}}{2^q t}\left\{c_2 \frac{p!\,(q+1)!}{(p+q+2)!} - c_3' \frac{(p+1)!\,(q+2)!}{2(p+q+4)!}\right\} + \mathcal{O}\left(\frac{1}{p^{q+4}}\right).$$

Moreover, $\frac{p!\,(q+1)!}{(p+q+2)!}$ and $\frac{(p+1)!\,(q+2)!}{(p+q+4)!}$ are both decreasing w.r.t p, so

$$|E(q,\infty,t) - E(q,p,t)| \leq |c_2| \frac{\mathrm{e}^{\alpha t}}{2^q t} \frac{p!\,(q+1)!}{(p+q+2)!} + \mathcal{O}\left(\frac{1}{p^{q+3}}\right).$$

■

Remark 5.5.5. *Whereas Proposition 5.5.2 in fact holds for any $h < 2\pi/t$, the proof of Proposition 5.5.3 is essentially based on the choice of $h = \pi/t$ since the key point is to be able to write $E(q,p+1,t) - E(q,p,t)$ as the general term of an alternating series. The impressive convergence rate of $E(q,p,t)$ definitely relies on the choice of this particular discretisation step. For a general step h, it is much more difficult to study the convergence rate and one cannot give an explicit upper-bound.*

Remark 5.5.6. *For $2\alpha t = 18.4$ and $q = p = 15$, the global error is bounded by $S_0 10^{-8} + t\,|f'(0) - \alpha f(0)|\,10^{-11}$. As one can see, the method we use to invert Laplace transforms provides a very good accuracy with few computations.*

Remark 5.5.7. *Considering the case of call options in Theorem 5.5.4 is sufficient since put prices are computed using parity relations and their accuracy is hung up to the one of call prices. Theorem 5.5.4 also holds for single barrier Parisian options.*

5.6 Numerical examples

In this section, we present some results obtained using the numerical inversion developed in Section 5.5. We have implemented our method in C and used the function `erfc` from the *Octave* library to compute the function $\mathcal{N}$ at a complex point. In the examples, we choose $p = 15$, $q = 15$ and $\alpha = 18.4/2T$. Hence, when the spot is of order 100 the accuracy of our method is ensured up to 10^{-6}.

In Table 5.1, we compare the prices of a double barrier Parisian out call with $S_0 = K = 100$, $L_1 = 90$, $L_2 = 110$, $r = 0.095$, $\delta = 0$ and $T = 1$ obtained with our method and the corrected Monte Carlo method of Baldi et al. [10] with 10000 samples. For the results obtained by the corrected Monte Carlo method, we precise the width of the confidence interval at level 95%. The accuracy showed by this approach decreases as the delay of the option increases. Our method is far more accurate and incredibly faster. For instance, if we consider the option described above with $D = 0.2$ and 250 time steps for the Monte Carlo, our algorithm takes 1.5 ms (CPU time) whereas the corrected Monte Carlo algorithm runs in 1.2 sec (CPU time).

Delay	MC Price	Price CI	Laplace
0.0500	0.545	0.0840	0.522
0.1000	1.142	0.1359	1.102
0.1500	1.774	0.1763	1.725
0.2000	2.241	0.2049	2.375
0.2500	3.044	0.2492	3.037
0.3000	3.681	0.2781	3.722
0.3500	4.530	0.3231	4.411
0.4000	4.933	0.3362	5.109

Table 5.1: Comparison corrected Monte Carlo and Laplace Transform

Figure 5.2 shows the evolution of the price of a double Parisian knock out call w.r.t. the delay when using the Laplace transform method or the corrected Monte Carlo one. We can see that the price given by the Laplace transform method is in the confidence interval given by the corrected Monte Carlo method. Figures 5.3 and 5.4 show the evolution of the price and the delta of a double barrier Parisian in call with respect to the spot and the strike. The delta is computed using a finite difference scheme.

5.7 Regularity of option prices

Proposition 5.7.1. *Let $f(t)$ be the "star" price of a double barrier Parisian option of maturity t. If $b_1 < 0$ and $b_2 > 0$, f is of class $\mathcal{C}^\infty$ and for all $k \geq 0$, $f^{(k)}(t) = \mathcal{O}\left(\mathrm{e}^{\frac{(m+\sigma)^2}{2}t}\right)$ when t goes to infinity.*

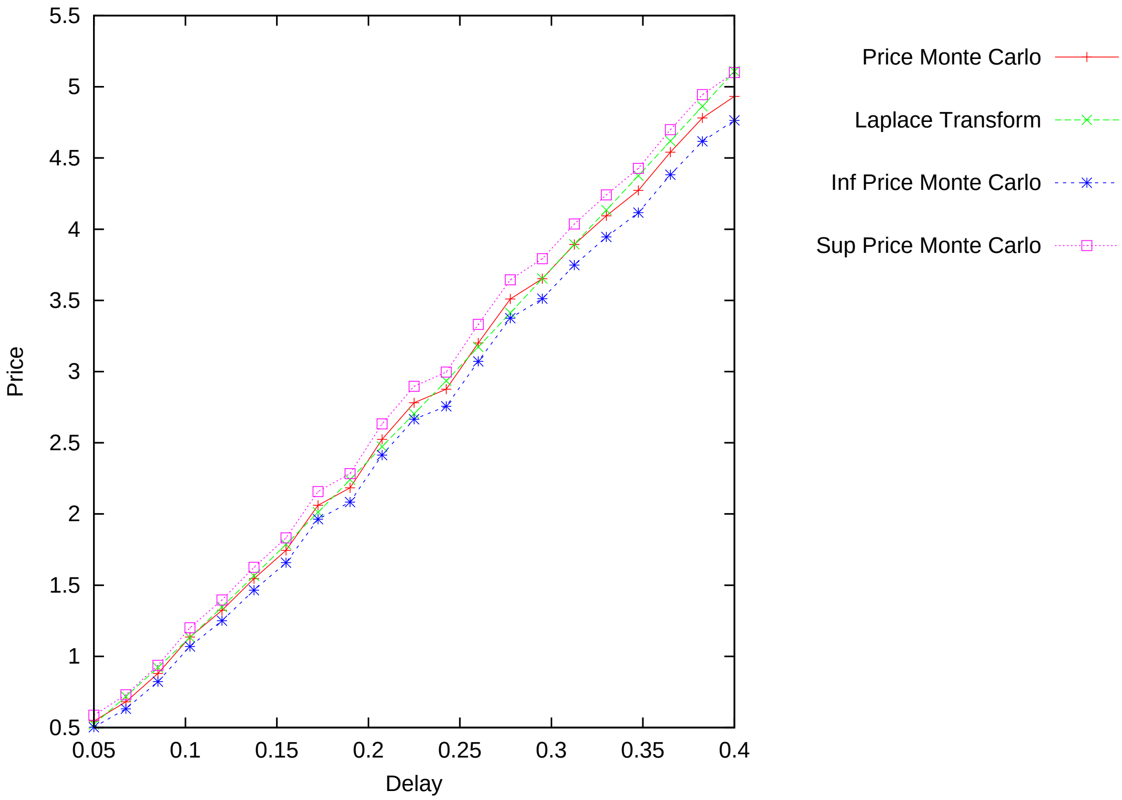

Figure 5.2: Comparison of corrected Monte Carlo and Laplace Transform

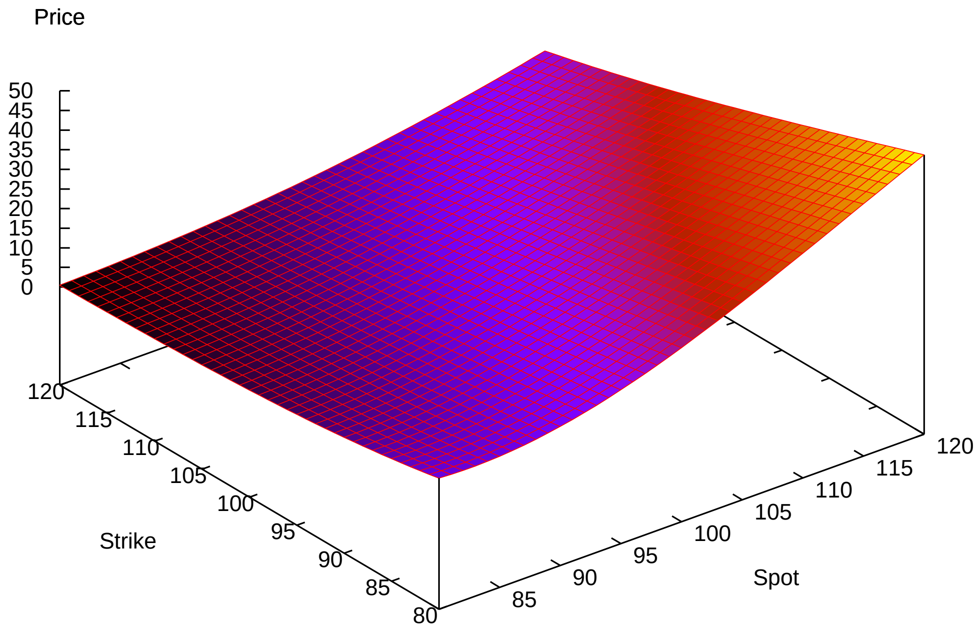

Figure 5.3: Price of a Double barrier Parisian In Call ($\sigma = 0.2$, $r = 0.02$, $\delta = 0$, $L = 80$, $U = 120$)

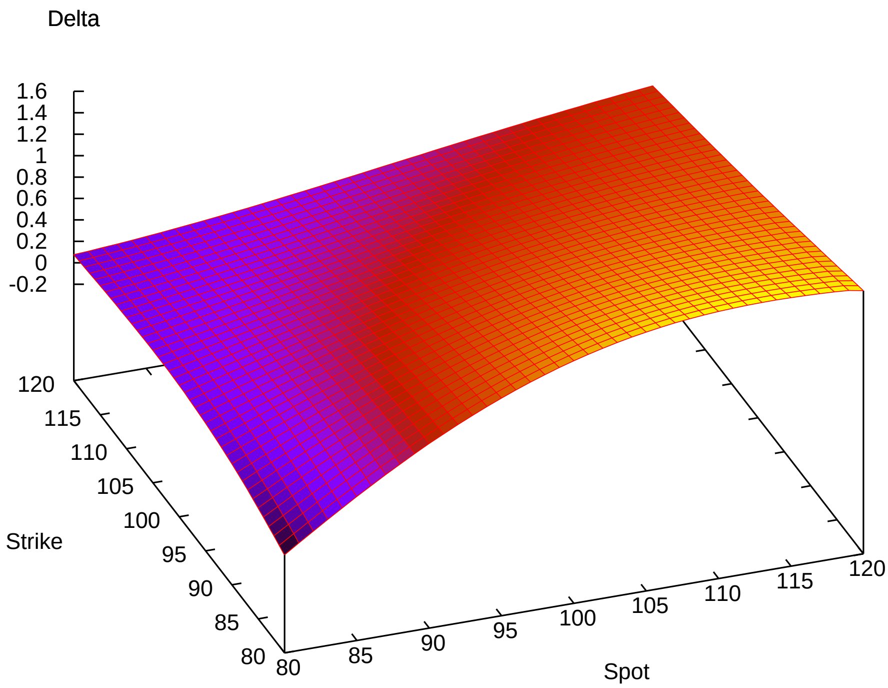

Figure 5.4: Delta of a Double barrier Parisian In Call

For the sake of clearness, we will only prove Proposition 5.7.1 for single barrier Parisian options as the scheme of the proof is still valid for double barrier Parisian options. Once again, we can restrict to calls. Let $f(t) = PDIC(x,t;K,L;r,\delta)$.

$$f(t) = \mathbb{E}\left[e^{mZ_t}(S_t - K)_+ \mathbf{1}_{\{T_b^- < t\}}\right].$$

Let W_t denote $Z_{t+T_b^-} - Z_{T_b^-}$. Relying on the strong Markov property,

$$f(t) = \mathbb{E}\left(\mathbf{1}_{\{T_b^- < t\}} \mathbb{E}\left[(xe^{\sigma(W_{t-\tau}+z)} - K)^+ e^{m(W_{t-\tau}+z)}\right]_{|z=Z_{T_b^-},\, \tau=T_b^-}\right). \tag{5.17}$$

Let ν denote the density of $Z_{T_b^-}$ (see Chesney et al. [23] for its expression) and μ the density of T_b^- (see Proposition 5.8.1 for a proof of existence). Since $Z_{T_b^-}$ and T_b^- are independent, Equation (5.17) can be written

$$f(t) = \int_0^t d\tau \int_{-\infty}^{\infty} dz \int_{-\infty}^{\infty} dw\ (xe^{\sigma(w\sqrt{t-\tau}+z)} - K)^+ e^{m(w\sqrt{t-\tau}+z)} p(w)\nu(z)\mu(\tau)$$

where $p(w) = \frac{1}{\sqrt{2\pi}}\,\mathrm{e}^{-\frac{w^2}{2}}$. A change of variable on τ gives

$$f(t) = \int_0^t d\tau \int_{-\infty}^{\infty} dz \int_{-\infty}^{\infty} dw\ (xe^{\sigma(w\sqrt{\tau}+z)} - K)^+ e^{m(w\sqrt{\tau}+z)} p(w)\nu(z)\mu(t-\tau).$$

Since μ is of class C^∞ and all its derivatives are null at 0 and bounded on any interval $[0,T]$(see Appendix 5.8), one can easily prove that f is of class C^∞ and that for all $k \geq 0$

$$f^{(k)}(t) = \int_0^t d\tau \int_{-\infty}^{\infty} dz \int_{-\infty}^{\infty} dw\ (xe^{\sigma(w\sqrt{\tau}+z)} - K)^+ e^{m(w\sqrt{\tau}+z)} p(w)\nu(z)\mu^{(k)}(t-\tau).$$

This proves the first part of Proposition 5.7.1. From Proposition 5.8.1, we know that μ and all its derivatives are bounded. Then, we can bound $f^{(k)}$

$$\begin{aligned}
\left|f^{(k)}(t)\right| &\leq \int_0^t d\tau \int_{-\infty}^{\infty} dz \int_{-\infty}^{\infty} dw\ xe^{(m+\sigma)(w\sqrt{\tau}+z)} p(w)\nu(z) \left\|\mu^{(k)}\right\|_\infty, \\
&\leq \int_{-\infty}^{\infty} xe^{(m+\sigma)z}\nu(z)dz \left\|\mu^{(k)}\right\|_\infty \int_0^t e^{\frac{(m+\sigma)^2}{2}\tau} d\tau, \\
&\leq e^{\frac{(m+\sigma)^2}{2}t} \frac{2x}{(m+\sigma)^2} \left\|\mu^{(k)}\right\|_\infty \int_{-\infty}^{\infty} e^{(m+\sigma)z}\nu(z)dz.
\end{aligned}$$

Relying on one more use of the strong Markov property, the same kind of computations can be reproduced for double barrier Parisian options.

5.8 Regularity of the density of T_b^-

In this section, we assume $b < 0$.

Proposition 5.8.1. *The r.v. T_b^- has a density μ w.r.t to Lebesgue's measure. μ is of class C^∞ and for all $k \geq 0$, $\mu^{(k)}(0) = \mu^{(k)}(\infty) = 0$.*

To prove this proposition, we need the two following lemmas.

Lemma 5.8.2. *Let $\mathcal{N}$ be the analytic prolongation of the cumulative normal distribution function on the complex plane. The following equivalent holds*

$$\mathcal{N}(r(1+i)) \sim 1 \text{ when } r \to \infty.$$

Lemma 5.8.3. *For $b < 0$, we have for $u \in \mathbb{R}$*

$$\mathbb{E}\left(\mathrm{e}^{-iuT_b^-}\right) = \mathcal{O}\left(\mathrm{e}^{-|b|\sqrt{|u|}}\right) \text{ when } |u| \to \infty.$$

Proof of Proposition 5.8.1. We recall that

$$\mathbb{E}\left(\mathrm{e}^{-\frac{\lambda^2}{2}T_b^-}\right) = \frac{\mathrm{e}^{\lambda b}}{\psi(\lambda\sqrt{D})}. \tag{5.18}$$

We define $\mathcal{O} = \{z \in \mathbb{C}; -\frac{\pi}{4} < \arg(z) < \frac{\pi}{4}\}$. One can easily prove that the function $z \longmapsto \mathbb{E}\left(\mathrm{e}^{-\frac{z^2}{2}T_b^-}\right)$ is holomorphic on the open set $\mathcal{O}$ and hence analytic. Moreover, $z \longmapsto \frac{\mathrm{e}^{zb}}{\psi(z\sqrt{D})}$ is also analytic on $\mathcal{O}$ except perhaps in a countable number of isolated points. These two functions coincide on $\mathbb{R}^+$, so they are equal on $\mathcal{O}$.

Consequently, we can derive the following equality. For all $z \in \mathbb{C}$ with positive real part, we have

$$\mathbb{E}\left(\mathrm{e}^{-zT_b^-}\right) = \frac{\mathrm{e}^{\sqrt{2z}b}}{\psi(\sqrt{2zD})}. \tag{5.19}$$

We use the following convention: for any $z \in \mathbb{C}$ with positive real part, $\sqrt{z}$ is the only complex number $z' \in \mathcal{O}$ such that $z = z'z'$.

Thanks to the continuity of both terms in (5.19), the equality also holds for pure imaginary numbers. Hence, by setting $z = iu$ for $u \in \mathbb{R}$ in Equation (5.19), we obtain the Fourier transform of T_b^-

$$\mathbb{E}\left(\mathrm{e}^{-iuT_b^-}\right) = \frac{\mathrm{e}^{\sqrt{2ui}b}}{\psi(\sqrt{2iuD})}.$$

From Lemma 5.8.3, we know that the Fourier transform of T_b^- is integrable on $\mathbb{R}$, thus the r.v. T_b^- has a density μ w.r.t. the Lebesgue measure given by

$$\mu(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \frac{\mathrm{e}^{\sqrt{2ui}b}}{\psi(\sqrt{2iuD})}\,\mathrm{e}^{-iut}\,du.$$

Moreover, thanks to Lemma 5.8.3, $u \longmapsto u^k \frac{e^{\sqrt{2uib}}}{\psi(\sqrt{2iuD})}$ is integrable and continuous. Hence, μ is of class C^∞. Since $\mu(t) = 0$ for $t < D$, for all $k \geq 0$, $\mu^{(k)}(0) = 0$. Lemma 5.8.4 yields that for all $k \geq 0$, $\lim_{t\to\infty} \mu^{(k)}(t) = 0$. ■

Proof of Lemma 5.8.2.

$$\mathcal{N}(x+iy) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{(v+iy)^2}{2}} \, dv.$$

It is easy to check that $\partial_x \mathcal{N}(x+iy) - \partial_y \mathcal{N}(x+iy) = 0$ and this definition coincides with the cumulative normal distribution function on the real axis, so it is the unique analytic prolongation. We write $\mathcal{N}(x+iy) = \mathcal{N}(x) + \int_0^y \partial_y \mathcal{N}(x+iy)$, to get

$$\begin{aligned} \mathcal{N}(x+iy) &= \mathcal{N}(x) - i\frac{1}{\sqrt{2\pi}} \int_0^y \int_{-\infty}^{x} (v+iu) \, e^{-\frac{(v+iu)^2}{2}} \, dv du, \\ &= \mathcal{N}(x) + i\frac{1}{\sqrt{2\pi}} \int_0^y e^{-\frac{(x+iu)^2}{2}} \, du. \end{aligned}$$

Taking $x + iy = r(1+i)$ gives

$$\begin{aligned} \mathcal{N}(r(1+i)) &= \mathcal{N}(r) + i\frac{1}{\sqrt{2\pi}} \int_0^r e^{-\frac{(r+iu)^2}{2}} \, du, \\ &= \mathcal{N}(r) + i\frac{1}{\sqrt{2\pi}} \int_0^1 e^{\frac{r^2}{2}(t^2-1)} \, e^{-itr^2} \, r dt. \end{aligned} \tag{5.20}$$

For $t \in [0,1)$, $e^{\frac{r^2}{2}(t^2-1)} \, r$ tends to 0 when r goes to infinity. The function $r \longmapsto e^{\frac{r^2}{2}(t^2-1)} \, r$ is maximum for $r = \frac{1}{1-t^2}$, hence the following upper bound holds

$$e^{\frac{r^2}{2}(t^2-1)} \, r \leq \frac{1}{1-t^2} \, e^{\frac{1}{2(t^2-1)}} \quad \text{for all } t \in [0,1).$$

The upper bound is integrable on $[0,1)$, so by using the bounded convergence theorem, we can assert that the integral on the right hand side of (5.20) tends to 0 when r goes to infinity. ■

Proof of Lemma 5.8.3. We only do the proof for $u > 0$. For $r > 0$,

$$\psi(r(1+i)) = 1 + r(1+i)\sqrt{2\pi} \, e^{r^2 i} \, \mathcal{N}(r(1+i)).$$

Using the equivalent of $\mathcal{N}(r(1+i))$ when r goes to infinity (see Lemma 5.8.2) enables to establish that $|\psi(r(1+i))| \sim 2r\sqrt{\pi}$ when r goes to infinity. Noticing that $\sqrt{iu} = \frac{\sqrt{2u}}{2}(1+i)$ ends the proof. ■

Here is a quite obvious lemma we used in the proof of Proposition 5.8.1.

Lemma 5.8.4. *Let g be an integrable function on $\mathbb{R}$, then*

$$\lim_{t\to\infty} \int_{-\infty}^{\infty} g(u) \, e^{iut} \, du = 0.$$

Annexe A

Quelques résultats bien connus

Nous rappelons ici quelques résultats que nous avons utilisés dans la preuve des Théorèmes 1.2.4 et 1.2.5.

Proposition A.0.5 (Tension et convergence en probabilité). *Soit $(X_n)_n$ une suite de v.a. dans $\mathbb{R}^d$ qui converge en probabilité vers zéro et $(Y_n)_n$ une suite de v.a. tendue dans $\mathbb{R}^d$. Alors, le produit scalaire $(\langle X_n | Y_n \rangle)_n$ converge en probabilité vers zéro.*

Proposition A.0.6 (Un résultat de convergence pour les intégrales). *Soit f une fonction continue par morceaux définie sur $\mathbb{R}^+$ telle que $f(u) \underset{u\to\infty}{\longrightarrow} l$, où l est un réel. Alors,*

$$\int_0^t e^{q(u-t)} f(u) du \underset{t\to\infty}{\longrightarrow} \frac{l}{q} \quad \forall t > 0, q > 0.$$

Intégrales et tension La proposition suivante est une extension de la Proposition A.0.6. Remarquons que le fait que le processus soit constant par morceaux est tout à fait essentiel puisqu'elle assure que les extrema courants sont atteints.

Proposition A.0.7. *Soit $(X(t))_{t\geq 0}$ un processus càdlàg constant par morceaux. Supposons que $X(t) \underset{t\to\infty}{\longrightarrow} x \in \mathbb{R}$ en probabilité. On définit $Y_t = \int_0^t e^{u-t} X_u du$. Alors, $Y(t) \xrightarrow[t\to\infty]{\mathbb{P}} x$.*

Démonstration. Puisque la convergence en loi vers une constante implique également la convergence en probabilité vers cette même constante, nous allons nous contenter de prouver la convergence en loi. Le première étape est de montrer que le processus $(Y_t)_t$ est tendu. Soit $M > 0$,

$$\begin{aligned}
\mathbb{P}(|Y_t| > M) &\leq \mathbb{P}\left(\int_0^t e^{u-t} |X_u| du > M\right), \\
&\leq \mathbb{P}\left(\sup_{u\in[0,t]} |X_u| > M\right). \qquad \text{(A.1)}
\end{aligned}$$

Puisque $(X(t))_{t\geq 0}$ est constant par morceaux, on a l'inclusion suivante

$$\left\{\sup_{u\in[0,t]}|X_u|\,;t\geq 0\right\}\subset\{X_s;s\geq 0\}\,.$$

Ce dernier ensemble est tendu puisque $X(\cdot)$ converge en probabilité. Par conséquent, la suite de variables aléatoires $(Y_t)_t$ est aussi tendue. La tension de $(Y_t)_t$ nous permet d'extraire une sous-suite convergente $(Y_{t_k})_k$. Notons L sa limite.

La deuxième étape consiste à montrer que $L=x$ presque sûrement. Quitte à procéder à une seconde extraction, on peut supposer que la suite $(t_k)_k$ est strictement croissante et peut-être choisie telle que $|t_{k+1}-t_k|\geq 1$ pour tout k. $\left(\int_0^{t_k}e^{u-t_k}(X_u-L)du\right)_k$ converge en loi vers zéro et donc en probabilité. La différence entre deux termes consécutifs de la suite tend vers zéro en probabilité.

$$\int_0^{t_k}(1-e^{t_k-t_{k+1}})e^{u-t_k}(X_u-L)du+\int_{t_k}^{t_{k+1}}e^{u-t_{k+1}}(X_u-L)du\xrightarrow[t\to\infty]{\mathbb{P}}0. \tag{A.2}$$

La première intégrale tend vers 0 en probabilité, par conséquent la seconde également. En utilisant la formule de la moyenne, on peut réécrire la seconde intégrale comme suit

$$\int_{t_k}^{t_{k+1}}e^{u-t_{k+1}}(X_u-L)du=\int_{t_k}^{t_{k+1}}e^{u-t_{k+1}}(c_k-L)du, \tag{A.3}$$

où

$$\inf_{u\in[t_k,t_{k+1}]}X_u\leq c_k\leq\sup_{u\in[t_k,t_{k+1}]}X_u. \tag{A.4}$$

De plus, puisque $|t_{k+1}-t_k|\geq 1$ pour tout k, la convergence en probabilité de l'intégrale implique que $c_k-L\xrightarrow[t\to\infty]{\mathbb{P}}0$. Comme le processus X est constant par morceaux, ces extrema sont atteints. Ainsi la convergence en probabilité de X_t vers x implique celle de $\inf_{u\in[t_k,t_{k+1}]}X_u$ et de $\sup_{u\in[t_k,t_{k+1}]}X_u$. La relation (A.4) nous permet d'affirmer que c_k converge en probabilité vers x. Ainsi, la convergence de c_k-L vers zéro en probabilité termine de montrer que $L=x$ presque sûrement.

Finalement, on a montré que toute sous-suite convergente de $(Y_t)_t$ converge vers x en loi. La suite $(Y_t)_t$ n'ayant qu'une seule valeur d'adhérence, elle converge vers x en loi. ■

TCL pour les martingales Nous rappelons ici un résultat sur la vitesse de convergence des tableaux triangulaires de martingales. Pour de plus amples détails sur le sujet nous renvoyons le lecteur au livre de Hall et Heyde [38]. Ici, nous nous contentons de rappeler le théorème suivant adapté de Duflo [29, Theorem 2.1.9].

Théorème A.0.8 (TCL pour les martingales)**.** *Supposons que $\{(\mathcal{F}_l^{(p)})_{0\leq l\leq p};p>0\}$ soit une famille de filtrations et que $\{(N_l^{(p)})_{0\leq l\leq p};p>0\}$ soit un tableau triangulaire de martingales de carré intégrable pour la filtration précédente. Supposons également que :*

($\mathcal{A}$A.1) *il existe une matrice symétrique définie positive V telle que $\langle N \rangle_p^{(p)}$ converge en probabilité vers V,*

($\mathcal{A}$A.2) *il existe $\rho > 0$ tel que*

$$\sum_{l=1}^{p} \mathbb{E}\left(\left\| N_l^{(p)} - N_{l-1}^{(p)} \right\|^{2+\rho} \middle| \mathcal{F}_{l-1}^{(p)} \right) \xrightarrow{\mathbb{P}} 0.$$

Alors

$$N_p^{(p)} \xrightarrow{\mathcal{L}} \mathcal{N}(0, V).$$

L'hypothèse ($\mathcal{A}$A.2) permet d'assurer que la condition de Lindeberg est vérifiée.

Appendix B

A few useful results for the Parisian options

In this chapter, we use the notations of Chapter 4. We recall here a few results used in Chapters 4 and 5. Sections B.1 and B.2 are devoted to the computations of two integrals appearing in the computation of the Laplace transforms of the Parisian option prices. In Section B.3, we recall a few important notions about the Brownian meander needed to compute the law of the couple $(T_b^-, Z_{T_b^-})$ as described in Section B.4.

B.1 The Laplace transform of μ_b in the case $b > 0$

We already know that $\mu_b(u) = \frac{|b|}{\sqrt{2\pi u^3}} e^{\left(\frac{-b^2}{2u}\right)}$. We use the notation $\theta = \sqrt{2\lambda}$.

$$\int_0^D e^{-\lambda u} \mu_b(du) = \int_0^D e^{-\frac{\theta^2}{2}u} \frac{b}{\sqrt{2\pi u^3}} e^{\frac{-b^2}{2u}} du.$$

The change of variable $t = \sqrt{\frac{b}{\theta}} \frac{1}{\sqrt{u}}$ leads to

$$\begin{aligned}
\int_0^D e^{-\lambda u} \mu_b(u) du &= \int_{\frac{\sqrt{b}}{\sqrt{\theta D}}}^{+\infty} \sqrt{\frac{2b\theta}{\pi}} \exp\left(\frac{-\theta b}{2}\left(\frac{1}{t^2} + t^2\right)\right) dt, \\
&= \int_{\frac{\sqrt{b}}{\sqrt{\theta D}}}^{+\infty} \sqrt{\frac{2b\theta}{\pi}} \exp\left(\frac{-\theta b}{2}\left(\frac{1}{t} - t\right)^2\right) e^{-\theta b} dt.
\end{aligned}$$

A new change of variable $v = \frac{1}{t} - t$ gives

$$\begin{aligned}
\int_0^D e^{-\lambda u} \mu_b(u) du &= \sqrt{\frac{b\theta}{2\pi}} e^{-\theta b} \int_{-\infty}^{\sqrt{\frac{\theta D}{b}} - \frac{\sqrt{b}}{\sqrt{\theta D}}} e^{\frac{-\theta b}{2} v^2} \left(1 - \frac{v}{\sqrt{v^2+4}}\right) dv, \\
&= \frac{1}{\sqrt{2\pi}} e^{-\theta b} \int_{-\infty}^{\theta\sqrt{D} - \frac{b}{\sqrt{D}}} e^{-u^2/2} \left(1 - \frac{u}{\sqrt{u^2 + 4\theta b}}\right) du.
\end{aligned}$$

A last change of variable $v = \sqrt{u^2 + 4\theta b}$ ends the computation

$$\hat{\mu_b}(\lambda) = e^{-\theta b}\mathcal{N}\left(\theta\sqrt{D} - \frac{b}{\sqrt{D}}\right) + e^{\theta b}\mathcal{N}\left(-\theta\sqrt{D} - \frac{b}{\sqrt{D}}\right).$$

If we let D go to infinity, we can deduce the Laplace transform of T_b, for any real b

$$\mathbb{E}[e^{-\lambda T_b}] = e^{-\sqrt{2\lambda}|b|}. \tag{B.1}$$

B.2 The valuation of $\int_0^{+\infty} \mathrm{e}^{-\lambda u} \frac{\mathrm{e}^{-\frac{x^2}{2u}}}{\sqrt{2\pi u}} du$

$$\int_0^{+\infty} \mathrm{e}^{-\lambda u} \frac{\mathrm{e}^{-\frac{x^2}{2u}}}{\sqrt{2\pi u}} du = \int_0^{+\infty} u\,\mathrm{e}^{-\lambda u} \frac{\mathrm{e}^{-\frac{x^2}{2u}}}{\sqrt{2\pi u^3}} du = \mathbb{E}(T_x\,\mathrm{e}^{-\lambda T_x}),$$

where T_x is the hitting time of level x of the Brownian Motion. We know that $\mathbb{E}(\mathrm{e}^{-\lambda T_x}) = \mathrm{e}^{-|x|\sqrt{2\lambda}}$ (see for instance Karatzas and Shreve [41, page 96]). The computation is ended by differentiating the above equatlity w.r.t λ. Doing so, we find Finally, we obtain

$$\int_0^{+\infty} e^{-\lambda u} \frac{\mathrm{e}^{-\frac{x^2}{2u}}}{\sqrt{2\pi u}} du = \frac{1}{\sqrt{2\lambda}} e^{-|x|\sqrt{2\lambda}}. \tag{B.2}$$

B.3 The Brownian meander

In this section, we only recall some useful results on the excursion theory of the Brownian motion. To find the proofs of the results announced, one can refer to Revuz and Yor [54] or Chung [24] for instance.

We denote by $g_t = \sup\{s \le t; Z_s = 0\}$ the left extremity of the excursion straddling time t. We define the slow Brownian filtration $(\mathcal{F}^+_{g_t}, t \ge 0)$ as $\mathcal{F}^+_{g_t} = \mathcal{F}^-_{g_t} \bigvee \sigma(\mathrm{sgn}(Z_t))$, where $\mathcal{F}^-_{g_t}$ is the $\sigma-$algebra generated by the random variables X_{g_t}, where X is a predictable process for the natural filtration of Brownian motion Z.

We denote by $g = g_1 = \sup\{s \le 1; Z_s = 0\}$ the left extremity of the excursion straddling time 1. The Brownian meander is defined as

$$m = \left\{ m_u = \frac{1}{\sqrt{1-g}} |Z_{g+u(1-g)}|;\ \ u \le 1 \right\}. \tag{B.3}$$

It is known that process m is independent of $\mathcal{F}^+_g$. The law of m_1 is given by

$$\mathbb{P}(m_1 \in dx) = x e^{-\frac{x^2}{2}}\, \mathbf{1}_{\{x>0\}}\, dx. \tag{B.4}$$

To find the law of m_1 we begin to calculate $\mathbb{P}(m_1 \le \lambda)$

$$\begin{aligned} \mathbb{P}(m_1 \le \lambda) &= \mathbb{P}(\tfrac{1}{\sqrt{1-g}}|Z_1| \le \lambda) \\ &= \mathbb{E}[\mathbf{1}_{\{\frac{1}{\sqrt{1-g}}|Z_1| \le \lambda\}}]. \end{aligned}$$

Thanks to Formula (B.21), we can write

$$\mathbb{P}(m_1 \leq \lambda) = \int_{s=0}^{+\infty} \int_{-\infty}^{+\infty} \mathbf{1}_{\{\frac{|x|}{\sqrt{1-s}} \leq \lambda\}} \mathbf{1}_{\{s \leq 1\}} \frac{|x|}{2\pi \sqrt{s(1-s)^3}} e^{-\frac{x^2}{2(1-s)}} dx ds.$$

For $\lambda \leq 0$, $\mathbb{P}(m_1 \leq \lambda) = 0$. From now on, we assume $\lambda \geq 0$. So, we get

$$\begin{aligned}
\mathbb{P}(m_1 \leq \lambda) &= \int_{s=0}^{1} \int_{x=-\lambda\sqrt{1-s}}^{+\lambda\sqrt{1-s}} \frac{|x|}{2\pi\sqrt{s(1-s)^3}} e^{-\frac{x^2}{2(1-s)}} dx ds, \\
&= 2\int_0^1 \int_0^{+\lambda\sqrt{1-s}} \frac{x}{2\pi\sqrt{s(1-s)^3}} e^{-\frac{x^2}{2(1-s)}} dx ds, \\
&= 2\int_0^1 \left[-\frac{1}{2\pi\sqrt{s(1-s)}} e^{-\frac{x^2}{2(1-s)}} \right]_{x=0}^{\lambda\sqrt{1-s}} ds, \\
&= 2\int_0^1 \frac{1}{2\pi\sqrt{s(1-s)}} (1 - e^{-\frac{\lambda^2}{2}}) ds, \\
&= \frac{(1 - e^{-\frac{\lambda^2}{2}})}{\pi} \int_0^1 \frac{1}{\sqrt{s(1-s)}} ds.
\end{aligned}$$

Now, we have to evaluate $\int_{s=0}^{1} \frac{1}{\sqrt{s(1-s)}} ds$. The change of variables $s = \cos^2\theta$ gives

$$\int_0^1 \frac{1}{\sqrt{s(1-s)}} ds = \int_{-\frac{\pi}{2}}^{0} -2\frac{\sin\theta\cos\theta}{|\sin\theta\cos\theta|} d\theta, = -2\int_{-\frac{\pi}{2}}^{0} \operatorname{sgn}(\sin(2\theta)) d\theta, = \pi.$$

So, we obtain

$$\mathbb{P}(m_1 \leq \lambda) = 1 - e^{-\frac{\lambda^2}{2}}.$$

For any $\lambda \in \mathbb{R}$, we have

$$\mathbb{P}(m_1 \in d\lambda) = \mathbf{1}_{\{\lambda \geq 0\}} \lambda e^{-\frac{\lambda^2}{2}} d\lambda. \tag{B.5}$$

Using Brownian scaling again, we can derive exactly the same results if we consider the excursion straddling time t instead of 1. Namely, we define

$$m^{(t)} = \left\{ m_u^{(t)} = \frac{1}{\sqrt{t - g_t}} |Z_{g_t + u(t-g_t)}|;\ u \leq 1 \right\}, \tag{B.6}$$

which is a Brownian meander independent of the σ-field $\mathcal{F}_{g_t}^{+}$. In particular, the law of $m^{(t)}$ does not depend on t.

Moreover, these results still hold if we consider a $\mathcal{F}$-stopping time instead of t. This last point is definitely essential as far as we are concerned and makes it possible to compute the law of $(T_b^-, Z_{T_b^-})$, as we do it in Appendix B.4.

B.3.1 The Azéma martingale

Now, we introduce the so-called Azéma martingale $\mu_t = \mathrm{sgn}(Z_t)\sqrt{(t-g_t)}$, which is a $\mathcal{F}_{g_t}^+$-martingale (see Azéma and Yor [9] for a detailed study of the Azéma martingale). We have

$$\mathbb{E}(\exp(\lambda Z_t - \frac{1}{2}\lambda^2 t)|\mathcal{F}_{g_t}^+) = \mathbb{E}(\exp(\lambda m_1^{(t)}\mu_t - \frac{1}{2}\lambda^2 t)|\mathcal{F}_{g_t}^+).$$

From the independence property we have just recalled, we get

$$\mathbb{E}(\exp(\lambda Z_t - \frac{1}{2}\lambda^2 t)|\mathcal{F}_{g_t}^+) = \exp(-\frac{1}{2}\lambda^2 t)\psi(\lambda\mu_t), \tag{B.7}$$

where

$$\psi(z) = \mathbb{E}(\exp(zm_1)) = \int_0^{+\infty} x\exp\left(zx - \frac{1}{2}x^2\right)dx. \tag{B.8}$$

B.4 The law of $(T_b^-, Z_{T_b^-})$ and $(T_b^+, Z_{T_b^+})$

We are interested in the law of $(T_b^-, Z_{T_b^-})$ and $(T_b^+, Z_{T_b^+})$. Such results can be found in Azéma and Yor [9], Revuz and Yor [54], Chung [24].

In the following, we consider a standard Brownian motion Z.

B.4.1 Case $b = 0$

In this case, we denote $T^- = T_0^-$. For any $t > 0$, we have

$$\{T^- \le t\} \;\; = \;\; \{\exists u \le t;\ Z_u \le 0 \text{ and } \sup_{s\in[u-D,u]} Z_u \le 0\}.$$

As Z is continuous, we can restrict to rational times

$$\{T^- \le t\} \;\; = \;\; \bigcup_{u\in\mathbb{Q},u\le t}\left(\{u-g_u \ge D\}\cap\{\mathrm{sgn}(Z_u) = -1\}\right).$$

Since each term composing the union belongs to $\mathcal{F}_t$, the random variable T^- is an $(\mathcal{F}_t)_t$ stopping time.

As recalled in Appendix B.3, we can use the definition of $m_u^{(t)}$ for $t = T^-$. Hence, the process $m^{(T^-)}$ defined by

$$m^{(T^-)} = \left\{m_u^{T^-} = \frac{1}{\sqrt{T^- - g_{T^-}}}|Z_{g_{T^-}+u(T^- - g_{T^-})}|; u \le 1\right\}$$

is a Brownian meander independent of $\mathcal{F}_{g_{T^-}}^+$.

As $g_{T^-} + D = T^-$, $\frac{1}{\sqrt{D}}Z_{T^-} = -m_1^{(T^-)}$, because Z_{T^-} is negative. Thus, Z_{T^-} is independent of $\mathcal{F}_{g_{T^-}}^+$ and we can deduce the law of Z_{T^-} from equation (B.5).

$$\mathbb{P}(Z_{T^-} \in dx) = -\frac{x}{D}\,\mathrm{e}^{-\frac{x^2}{2D}}\,\mathbf{1}_{\{x<0\}}dx. \tag{B.9}$$

Moreover, as T^- is $\mathcal{F}^+_{g_{T^-}}$-measurable, it comes out that Z_{T^-} and T^- are independent.

Using Equation (B.7), the process $\{\psi(-\lambda\mu_t)\,\mathrm{e}^{-\frac{1}{2}\lambda^2 t}, t \geq 0\}$ is a $\mathcal{F}^+_{g_t}$ martingale for any $\lambda > 0$. By using some localisation technique, we can apply the optional stopping time theorem at T^- to obtain

$$\mathbb{E}(\psi(-\lambda\mu_{T^-})\,\mathrm{e}^{-\frac{1}{2}\lambda^2 T^-}) = \psi(0) = 1.$$

Since $\mu_{T^-} = -\sqrt{D}$, we get

$$\mathbb{E}\left(\mathrm{e}^{-\frac{1}{2}\lambda^2 T^-}\right) = \frac{1}{\psi(\lambda\sqrt{D})}. \tag{B.10}$$

Similarly, $\frac{1}{\sqrt{D}} Z_{T^+} = m_1^{T^-}$ and the law of Z_{T^+} is given by

$$\mathbb{P}(Z_{T^+} \in dx) = \frac{x}{D}\,\mathrm{e}^{-\frac{x^2}{2D}}\,\mathbf{1}_{\{x>0\}}dx. \tag{B.11}$$

With exactly the same method, we find

$$\mathbb{E}\left(\mathrm{e}^{-\frac{1}{2}\lambda^2 T^+}\right) = \frac{1}{\psi(\lambda\sqrt{D})}. \tag{B.12}$$

B.4.2 Case $b < 0$

This case can be reduced to the previous one with the help of the stopping time T_b. We can write $T_b^- = T_b + T^-(W)$, with

$$\begin{array}{rcl} T_0^-(W) & = & \inf\{t \geq 0; \mathbf{1}_{\{W_t \leq 0\}}(t - g_t^W) \geq D\} \stackrel{law}{=} T_0^-, \\ W & = & \{W_t = Z_{T_b+t} - b; t \geq 0\}, \\ g_t^W & = & \sup\{u \leq t; W_u = 0\}. \end{array} \tag{B.13}$$

Moreover, using the strong Markov property, it is clear that T_b and $T_0^-(W)$ are independent.

$$\mathbb{E}(\mathrm{e}^{-\frac{1}{2}\lambda^2 T_b^-}) = \mathbb{E}(\mathrm{e}^{-\frac{1}{2}\lambda^2 T_b})\mathbb{E}(\mathrm{e}^{-\frac{1}{2}\lambda^2 T_0^-(W)}).$$

As $\mathbb{E}(\exp(-\frac{1}{2}\lambda^2 T_b)) = \exp(-|b|\lambda)$, we get

$$\mathbb{E}\left(\mathrm{e}^{-\frac{1}{2}\lambda^2 T_b^-}\right) = \frac{\mathrm{e}^{b\lambda}}{\psi(\lambda\sqrt{D})}. \tag{B.14}$$

Now, we are trying to find the law of $Z_{T_b^-}$

$$\mathbb{P}(Z_{T_b^-} \in dx) = \mathbb{E}(\mathbb{E}(\mathbf{1}_{\{Z_{T_b^-} \in dx\}}|\mathcal{F}_{T_b})).$$

Using the definition of W (see (B.13)), we get

$$\mathbb{P}(Z_{T_b^-} \in dx) = \mathbb{E}(\mathbf{1}_{\{W_{T_b^-(Z)-T_b(Z)} \in dx-b\}}) = \mathbb{E}(\mathbf{1}_{\{W_{T_0^-} \in dx-b\}}).$$

Finally, we obtain

$$\mu^-(dx) = \mathbb{P}(Z_{T_b^-} \in dx) = -\frac{x}{D}\,\mathrm{e}^{-\frac{(x-b)^2}{2D}}\,\mathbf{1}_{\{x<b\}}dx. \tag{B.15}$$

B.4.3 Case $b > 0$

Using the same tricks as before, we find

$$\mathbb{E}\left(\mathrm{e}^{-\frac{1}{2}\lambda^2 T_b^-}\right) = \frac{\mathrm{e}^{-b\lambda}}{\psi(\lambda\sqrt{D})}. \tag{B.16}$$

$$\mu^+(dx) = \mathbb{P}(Z_{T_b^+} \in dx) = -\frac{x-b}{D}\,\mathrm{e}^{-\frac{(x-b)^2}{2D}}\,\mathbf{1}_{\{x>b\}}dx. \tag{B.17}$$

B.5 Laplace transforms of Parisian times

B.5.1 Laplace transforms for T_b^-, T_b^+, $Z_{T_b^-}$ and $Z_{T_b^+}$

The following Lemma ensues from Chesney et al. [23, Sections 8.3 and 8.4], giving the Laplace transforms of T_b^- and T_b^+, for $b \in \mathbb{R}$.

Lemma B.5.1. *Let θ denote $\sqrt{2\lambda}$. We recall that the function ψ is given in Definition B.8. Then,*

$$\begin{aligned}
\mathbb{E}[e^{-\lambda T_b^-}] &= \frac{e^{\theta b}}{\psi(\theta\sqrt{D})} \qquad \text{for } b<0,\\
&= e^{-\lambda D}\left(1 - 2\mathcal{N}\left(-\frac{b}{\sqrt{D}}\right)\right)\\
&\quad + \psi(\theta\sqrt{D})^{-1}\left[e^{-\theta b}\mathcal{N}\left(\theta\sqrt{D} - \frac{b}{\sqrt{D}}\right) + e^{\theta b}\mathcal{N}\left(-\theta\sqrt{D} - \frac{b}{\sqrt{D}}\right)\right] \ \text{otherwise.}
\end{aligned}$$

$$\begin{aligned}
\mathbb{E}[e^{-\lambda T_b^+}] &= \frac{e^{-\theta b}}{\psi(\theta\sqrt{D})} \qquad \text{for } b>0,\\
&= e^{-\lambda D}\left(1 - 2\mathcal{N}\left(\frac{b}{\sqrt{D}}\right)\right)\\
&\quad + \psi(\theta\sqrt{D})^{-1}\left[e^{\theta b}\mathcal{N}\left(\theta\sqrt{D} + \frac{b}{\sqrt{D}}\right) + e^{-\theta b}\mathcal{N}\left(-\theta\sqrt{D} + \frac{b}{\sqrt{D}}\right)\right] \ \text{otherwise.}
\end{aligned}$$

To compute the Laplace transforms of $Z_{T_b^-}$ and $Z_{T_b^+}$ we use the densities of these random variables. We refer to Chesney et al. [23, Sections 8.3 and 8.4] for an expression of these densities (note that there is a misprint in Chesney et al. [23, Appendix 8.3]: indicator $\mathbf{1}_{\{x<b\}}$ is missing in the formula of $\mathbb{P}(Z_D \in dx, T_b > D)$ case $b > 0$).

Lemma B.5.2. *Let θ denote $\sqrt{2\lambda}$. We recall that the function ψ is given in Definition B.8. Then,*

$$\begin{aligned}\mathbb{E}\left[e^{-\theta Z_{T_b^-}}\right] &= e^{-\theta b}\psi(\theta\sqrt{D}) \quad \text{for } b < 0,\\ &= 2\mathcal{N}\left(-\frac{b}{\sqrt{D}}\right) e^{-\theta b}\psi(\theta\sqrt{D})\\ &+ e^{\lambda D}\left[\mathcal{N}\left(\frac{b}{\sqrt{D}} + \theta\sqrt{D}\right) - e^{-2\theta b}\mathcal{N}\left(-\frac{b}{\sqrt{D}} + \theta\sqrt{D}\right)\right] \text{ otherwise.}\end{aligned}$$

$$\begin{aligned}\mathbb{E}\left[e^{-\theta Z_{T_b^+}}\right] &= e^{-\theta b}\psi(-\theta\sqrt{D}) \quad \text{for } b > 0,\\ &= 2\mathcal{N}\left(\frac{b}{\sqrt{D}}\right) e^{-\theta b}\psi(-\theta\sqrt{D})\\ &+ e^{\lambda D}\left[\mathcal{N}\left(-\frac{b}{\sqrt{D}} - \theta\sqrt{D}\right) - e^{-2\theta b}\mathcal{N}\left(\frac{b}{\sqrt{D}} - \theta\sqrt{D}\right)\right] \text{ otherwise.}\end{aligned}$$

B.5.2 Formulas for $\mathbb{E}[e^{-\lambda T_{b_1}^-}\mathbf{1}_{\{T_{b_1}^- < T_{b_2}^+\}}]$ and $\mathbb{E}[e^{-\lambda T_{b_2}^+}\mathbf{1}_{\{T_{b_2}^+ < T_{b_1}^-\}}]$

Lemma B.5.3. *Let θ denote $\sqrt{2\lambda}$. We recall that the function ψ is given in Definition B.8. Then,*

$$\begin{aligned}\mathbb{E}[e^{-\lambda T_{b_1}^-}\mathbf{1}_{\{T_{b_1}^- < T_{b_2}^+\}}] &= \frac{\mathbb{E}[e^{-\lambda T_{b_1}^-}] - a_1\mathbb{E}[e^{-\lambda T_{b_2}^+}]}{1 - a_1 a_2},\\ \mathbb{E}[e^{-\lambda T_{b_2}^+}\mathbf{1}_{\{T_{b_2}^+ < T_{b_1}^-\}}] &= \frac{\mathbb{E}[e^{-\lambda T_{b_2}^+}] - a_2\mathbb{E}[e^{-\lambda T_{b_1}^-}]}{1 - a_1 a_2},\end{aligned}$$

where $a_1 = \frac{e^{\theta b_1}}{\psi(\theta\sqrt{D})}\mathbb{E}[\exp(-\theta Z_{T_{b_2}^+})]$ and $a_2 = \frac{e^{-\theta b_2}}{\psi(\theta\sqrt{D})}\mathbb{E}[\exp(\theta Z_{T_{b_1}^-})]$.

Proof. Let E_- (resp. E_+) denote $\mathbb{E}[e^{-\lambda T_{b_1}^-}\mathbf{1}_{\{T_{b_1}^- < T_{b_2}^+\}}]$ (resp. $\mathbb{E}[e^{-\lambda T_{b_2}^+}\mathbf{1}_{\{T_{b_2}^+ < T_{b_1}^-\}}]$). Then,

$$\begin{aligned}\mathbb{E}[e^{-\lambda T_{b_1}^-}] &= E_- + \mathbb{E}[e^{-\lambda T_{b_1}^-}\mathbf{1}_{\{T_{b_2}^+ < T_{b_1}^-\}}],\\ \mathbb{E}[e^{-\lambda T_{b_2}^+}] &= E_+ + \mathbb{E}[e^{-\lambda T_{b_2}^+}\mathbf{1}_{\{T_{b_1}^- < T_{b_2}^+\}}].\end{aligned}$$

We can write $\mathbb{E}[e^{-\lambda T_{b_1}^-}\mathbf{1}_{\{T_{b_2}^+<T_{b_1}^-\}}] = \mathbb{E}[\mathbb{E}[\exp(-\lambda(T_{b_2}^+ + \tilde{T}_{b_1 - Z_{T_{b_2}^+}}^-))\big|\mathcal{F}_{T_{b_2}^+}]\mathbf{1}_{\{T_{b_2}^+<T_{b_1}^-\}}]$, where $\tilde{T}$ denotes a stopping time related to a Brownian motion independent of $\mathcal{F}_{T_{b_2}^+}$. Hence, we get $\mathbb{E}[e^{-\lambda T_{b_1}^-}\mathbf{1}_{\{T_{b_2}^+<T_{b_1}^-\}}] = \mathbb{E}[\exp(-\lambda T_{b_2}^+)\mathbf{1}_{\{T_{b_2}^+<T_{b_1}^-\}}\mathbb{E}[\exp(-\lambda(\tilde{T}_{b_1-z}^-))]_{|z=Z_{T_{b_2}^+}}]$. Since $b_1 - Z_{T_{b_2}^+} < 0$, we get

$$\mathbb{E}[e^{-\lambda T_{b_1}^-}\mathbf{1}_{\{T_{b_2}^+<T_{b_1}^-\}}] = \frac{e^{\theta b_1}}{\psi(\theta\sqrt{D})}\mathbb{E}\left[\exp(-\lambda T_{b_2}^+)\mathbf{1}_{\{T_{b_2}^+<T_{b_1}^-\}}e^{-\theta Z_{T_{b_2}^+}}\right].$$

From Chesney et al. [23, Sections 8.3 and 8.4] we know that $T_{b_2}^+$ is an $\mathcal{F}_{g_t}^+$ stopping time whereas $Z_{T_{b_2}^+}$ is independent of $\mathcal{F}_{g_{T_{b_2}^+}}^+$. Hence, we get $\mathbb{E}[e^{-\lambda T_{b_1}^-}\mathbf{1}_{\{T_{b_2}^+<T_{b_1}^-\}}] = \frac{e^{\theta b_1}}{\psi(\theta\sqrt{D})}\mathbb{E}[e^{-\theta Z_{T_{b_2}^+}}] = a_1 E_+$. The same kind of proof leads to $\mathbb{E}[e^{-\lambda T_{b_2}^+}\mathbf{1}_{\{T_{b_1}^-<T_{b_2}^+\}}] = a_2 E_-$, which ends the proof. ■

B.6 Around Brownian Motion

Let us consider a standard Brownian motion $W = \{W_t; t \geq 0\}$. First of all, we recall two results on the joint law of the Brownian motion and its extrema. A proof can be found in Revuz and Yor [54].

Law of $(W_t, \sup_{0\leq u\leq t} W_u)$

$$\mathbb{P}(W_t \in dx, \sup_{0\leq u\leq t} W_u \in dy) = \mathbf{1}_{\{0\leq y\}}\mathbf{1}_{\{x\leq y\}}\frac{2(2y-x)}{\sqrt{2\pi t^3}}\,\mathrm{e}^{-\frac{(2y-x)^2}{2t}}\,dxdy. \tag{B.18}$$

Law of $(W_t, \inf_{0\leq u\leq t} W_u)$

$$\mathbb{P}(W_t \in dx, \inf_{0\leq u\leq t} W_u \in dy) = \mathbf{1}_{\{y\leq 0\}}\mathbf{1}_{\{y\leq x\}}\frac{2(2y-x)}{\sqrt{2\pi t^3}}\,\mathrm{e}^{-\frac{(2y-x)^2}{2t}}\,dxdy \tag{B.19}$$

Hitting time The purpose is to find the law of T_b defined as

$$T_b = \inf\{t \geq 0 \mid W_t = b\}.$$

Using the law of the running infimum and supremum of the Brownian motion (see Equations (B.18) and B.19), it is easy to prove that for any $b \in \mathbb{R}$,

$$\mathbb{P}(T_b \in dx) = \frac{|b|}{\sqrt{2\pi x^3}}e^{-\frac{b^2}{2x}}dx. \tag{B.20}$$

Excursion Let g_t denote the last time before t that W hit the level 0.

$$g_t = \sup\{u \leq t \mid W_u = 0\}.$$

$$\mathbb{P}(W_t \in dy, g_t \in ds) = \frac{|y|}{2\pi\sqrt{s(t-s)^3}} \mathrm{e}^{-\frac{y^2}{2(t-s)}} \mathbf{1}_{\{s \leq t\}} \, ds \, dy. \tag{B.21}$$

Bibliographie

[1] J. Abate et W. Whitt. Numerical inversion of Laplace transforms of probability distributions. *ORSA Journal of Computing*, 7(1), Winter 1995.

[2] J. Abate, L. Choudhury, et G. Whitt. An introduction to numerical transform inversion and its application to probability models. *Computing Probability*, pages 257 – 323, 1999.

[3] J. Anderluh et H. van der Weide. Parisian options—the implied barrier concept. In *Computational science—ICCS 2004. Part IV*, volume 3039 of *Lecture Notes in Comput. Sci.*, pages 851–858. Springer, Berlin, 2004.

[4] L. Andersen et R. Brotherton-Ratcliffe. Exact exotics. *Risk*, 9(10) :85–89, 1996.

[5] C. Andrieu, É. Moulines, et P. Priouret. Stability of stochastic approximation under verifiable conditions. *SIAM J. Control Optim.*, 44(1) :283–312 (electronic), 2005. ISSN 0363-0129.

[6] B. Arouna. Adaptative Monte Carlo method, a variance reduction technique. *Monte Carlo Methods Appl.*, 10(1) :1–24, 2004. ISSN 0929-9629.

[7] B. Arouna. Robbins-Monro algorithms and variance reduction in finance. *The Journal of Computational Finance*, 7(2), Winter 2003/2004.

[8] M. Avellaneda et L. Wu. Pricing Parisian-style options with a lattice method. *International Journal of Theoretical and Applied Finance*, 2(1) :1–16, 1999.

[9] J. Azéma et M. Yor. Étude d'une martingale remarquable. In *Séminaire de Probabilités, XXIII*, volume 1372 of *Lecture Notes in Math.*, pages 88–130. Springer, Berlin, 1989.

[10] P. Baldi, L. Caramellino, et M. G. Iovino. Pricing complex barrier options with general features using sharp large deviation estimates. In *Monte Carlo and quasi-Monte Carlo methods 1998 (Claremont, CA)*, pages 149–162. Springer, Berlin, 2000.

[11] M. Benaïm. A dynamical system approach to stochastic approximations. *SIAM J. Control Optim.*, 34(2) :437–472, 1996. ISSN 0363-0129.

[12] M. Benaïm. Dynamics of stochastic approximation algorithms. In *Séminaire de Probabilités, XXXIII*, volume 1709 of *Lecture Notes in Math.*, pages 1–68. Springer, Berlin, 1999.

[13] A. Benveniste, M. Métivier, et P. Priouret. *Adaptive algorithms and stochastic approximations*, volume 22 of *Applications of Mathematics (New York)*. Springer-Verlag, Berlin, 1990. ISBN 3-540-52894-6. Translated from the French by Stephen S. Wilson.

[14] P. Billingsley. *Convergence of Probability Measures*. Wiley, New York, 1968.

[15] F. Black et M. Scholes. The pricing of options and corporate liabilities. *Journal of Political Economy*, 81 :637–654, 1973.

[16] J. F. Bonnans, J. C. Gilbert, C. Lemaréchal, et C. A. Sagastizábal. *Numerical optimization*. Universitext. Springer-Verlag, Berlin, second edition, 2006. ISBN 3-540-35445-X. Theoretical and practical aspects.

[17] V. S. Borkar et S. P. Meyn. The O.D.E. method for convergence of stochastic approximation and reinforcement learning. *SIAM J. Control Optim.*, 38(2) :447–469 (electronic), 2000. ISSN 0363-0129.

[18] C. Bouton. *Approximation Gaussienne d'algorithmes stochastiques à dynamique Markovienne*. PhD thesis, Université Pierre et Marie Curie - Paris 6, 1985.

[19] R. Buche et H. J. Kushner. Rate of convergence for constrained stochastic approximation algorithms. *SIAM J. Control Optim.*, 40(4) :1011–1041 (electronic), 2001/02.

[20] J. Chancelier, B. Pinçon, et F. Delbecque. Nsp. `http://cermics.enpc.fr/~jpc/scilab/site/Scilab-Gtk/Scilab-Gtk/index.htm`.

[21] H. Chen et Y. Zhu. *Stochastic Approximation Procedure with randomly varying truncations*. Scientia Sinica Series, 1986.

[22] H. F. Chen, G. Lei, et A. J. Gao. Convergence and robustness of the Robbins-Monro algorithm truncated at randomly varying bounds. *Stochastic Process. Appl.*, 27(2) : 217–231, 1988. ISSN 0304-4149.

[23] M. Chesney, M. Jeanblanc-Picqué, et M. Yor. Brownian excursions and Parisian barrier options. *Adv. in Appl. Probab.*, 29(1) :165–184, 1997. ISSN 0001-8678.

[24] K. L. Chung. Excursions in Brownian motion. *Ark. Mat.*, 14(2) :155–177, 1976. ISSN 0004-2080.

[25] M. Costabile. A combinatorial approach for pricing Parisian options. *Decis. Econ. Finance*, 25(2) :111–125, 2002. ISSN 1593-8883.

[26] J. Cox, S. Ross, et M. Rubinstein. Option pricing : A simplified approach. *Journal of Financial Economics*, (7) :229–263, 1979.

[27] B. Delyon. Stochastic approximation with decreasing gain : Convergence and asymptotic theory. Unpublished lecture notes, Université de Rennes, 2000.

[28] B. Delyon. General results on the convergence of stochastic algorithms. *IEEE Transactions on Automatic Control*, 41(9) :1245–1255, 1996.

[29] M. Duflo. *Random Iterative Models.* Springer-Verlag Berlin and New York, 1997.

[30] P. Dupuis et H. J. Kushner. Asymptotic behavior of constrained stochastic approximations via the theory of large deviations. *Probab. Theory Related Fields*, 75 (2) :223–244, 1987.

[31] J.-C. Fort et G. Pagès. Convergence of stochastic algorithms : from the Kushner-Clark theorem to the Lyapounov functional method. *Adv. in Appl. Probab.*, 28(4) : 1072–1094, 1996. ISSN 0001-8678.

[32] J.-C. Fort et G. Pagès. Asymptotic behavior of a Markovian stochastic algorithm with constant step. *SIAM J. Control Optim.*, 37(5) :1456–1482 (electronic), 1999. ISSN 0363-0129.

[33] M. Gahbiche et M. Pelletier. On the estimation of the asymptotic covariance matrix for the averaged Robbins-Monro algorithm. *C. R. Acad. Sci. Paris Sér. I Math.*, 331(3) :255–260, 2000. ISSN 0764-4442.

[34] P. Glasserman. *Monte Carlo methods in financial engineering*, volume 53 of *Applications of Mathematics (New York).* Springer-Verlag, New York, 2004. ISBN 0-387-00451-3. , Stochastic Modelling and Applied Probability.

[35] P. Glynn et M. Hsieh. Confidence regions for stochastic approximation algorithms. In *2002 Winter simulation Conference*, 2002.

[36] R. Haber, P. Schonbucher, et P.Wilmott. An American in Paris. OFRC Working Papers Series 1999mf14, Oxford Financial Research Centre, 1999. available at `http://ideas.repec.org/p/sbs/wpsefe/1999mf14.html`.

[37] R. Haber, P. Schonbucker, et P. Wilmott. Pricing Parisian options. *Journal of Derivatives*, (6) :71–79, 1999.

[38] P. Hall et C. C. Heyde. *Martingale limit theory and its application.* Academic Press Inc. [Harcourt Brace Jovanovich Publishers], New York, 1980. ISBN 0-12-319350-8. Probability and Mathematical Statistics.

[39] P. Hartley. Pricing Parisian options by Laplace inversion. *Decisions in Economics & Finance*, 2002.

[40] J. Jacod et A. Shiryaev. *Limit Theorems for Stochastic Processes*. Springer-Verlag Berlin, 1987.

[41] I. Karatzas et S. E. Shreve. *Brownian motion and stochastic calculus*, volume 113 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, second edition, 1991. ISBN 0-387-97655-8.

[42] H. J. Kushner et D. S. Clark. *Stochastic approximation methods for constrained and unconstrained systems*, volume 26 of *Applied Mathematical Sciences*. Springer-Verlag, New York, 1978. ISBN 0-387-90341-0.

[43] H. J. Kushner et J. Yang. Stochastic approximation with averaging of the iterates : Optimal asymptotic rate of convergence for general processes. *SIAM Journal on Control and Optimization*, 31(4) :1045–1062, 1993.

[44] H. J. Kushner et G. G. Yin. *Stochastic approximation and recursive algorithms and applications*, volume 35 of *Applications of Mathematics (New York)*. Springer-Verlag, New York, second edition, 2003. ISBN 0-387-00894-2. Stochastic Modelling and Applied Probability.

[45] C. Labart et J. Lelong. Pricing Parisian options. Technical report, ENPC, `http://cermics.enpc.fr/reports/CERMICS-2005/CERMICS-2005-294.pdf`, December 2005.

[46] C. Labart et J. Lelong. Pricing double barrier Parisian options using Laplace transforms. Technical report, ENPC, `http://cermics.enpc.fr/reports/CERMICS-2006/CERMICS-2006-328.pdf`, November 2006.

[47] D. Lamberton et B. Lapeyre. *Introduction to stochastic calculus applied to finance*. Chapman & Hall, London, 1996. ISBN 0-412-71800-6. Translated from the 1991 French original by Nicolas Rabeau and François Mantion.

[48] A. Mokkadem et M. Pelletier. Convergence rate and averaging of nonlinear two-time-scale stochastic approximation algorithms. *Ann. Appl. Probab.*, 16(3) :1671–1702, 2006. ISSN 1050-5164.

[49] M. Pelletier. Weak convergence rates for stochastic approximation with application to multiple targets and simulated annealing. *Ann. Appl. Probab.*, 8(1) :10–44, 1998. ISSN 1050-5164.

[50] B. T. Polyak et A. B. Juditsky. Acceleration of stochastic approximation by averaging. *SIAM J. Control Optim.*, 30(4) :838–855, 1992. ISSN 0363-0129.

[51] PREMIA. Option pricer developped by the mathfi tream, INRIA, Rocquencourt. FRANCE.

[52] P. Protter. *Stochastic integration and Differential Equations.* Springer-Verlag Berlin, second edition, 2004.

[53] F. Quittard-Pinon, C. Bernard, et O. Le Courtois. Evaluation numérique d'options parisiennes. *Banque & Marchés*, 2004.

[54] D. Revuz et M. Yor. *Continuous martingales and Brownian motion*, volume 293 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences].* Springer-Verlag, Berlin, third edition, 1999. ISBN 3-540-64325-7.

[55] H. Robbins et S. Monro. A stochastic approximation method. *Ann. Math. Statistics*, 22 :400–407, 1951.

[56] L. Rogers et D. Williams. *Diffusions, Markov Processes and Martingales*, volume 2 (Itô Calculus). Cambridge University Press, 2000.

[57] M. Schröder. Brownian excursions and Parisian barrier options : a note. *J. Appl. Probab.*, 40(4) :855–864, 2003. ISSN 0021-9002.

[58] D. V. Widder. *The Laplace Transform.* Princeton Mathematical Series, v. 6. Princeton University Press, Princeton, N. J., 1941.

[59] P. Wilmott. *Derivatives.* University Edition, 1998.